

5808 #4

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶:</p> <p>C12N 15/55, 9/16, 1/15, 1/21, A23K 1/165, C12P 19/14 // (C12N 1/15, C12R 1:77) (C12N 1/21, C12R 1:19)</p>	<p>A1</p>	<p>(11) International Publication Number: WO 97/35016</p> <p>(43) International Publication Date: 25 September 1997 (25.09.97)</p>
<p>(21) International Application Number: PCT/US97/03866</p> <p>(22) International Filing Date: 10 March 1997 (10.03.97)</p> <p>(30) Priority Data: 08/617,235 18 March 1996 (18.03.96) US</p>		<p>(81) Designated States: AL, AU, BB, BG, BR, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p>
<p>(71) Applicant: NOVO NORDISK BIOTECH, INC. [US/US]; 1445 Drew Avenue, Davis, CA 95616-4880 (US).</p> <p>(72) Inventors: BERKA, Randy, M.; 3609 Modoc, Davis, CA 95616 (US). RAY, Michael, W.; 605 Robin Place, Davis, CA 95616 (US). KLOTZ, Alan, V.; 725 Collier Drive, Dixon, CA 95620 (US).</p> <p>(74) Agents: ZELSON, Steve, T. et al.; Novo Nordisk of North America, Inc., Suite 6400, 405 Lexington Avenue, New York, NY 10174 (US).</p>		

(54) Title: POLYPEPTIDES HAVING PHYTASE ACTIVITY AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract

The present invention relates to isolated polypeptides having phytase activity from *Thermomyces lanuginosus* and isolated nucleic acid sequences encoding the polypeptides. The invention also relates to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences as well as methods for producing the polypeptides. The present invention further relates to composite feeds and methods of reducing phytate levels.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bohnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

POLYPEPTIDES HAVING PHYTASE ACTIVITY AND NUCLEIC ACIDS ENCODING SAME

Background of the Invention

5

Field of the Invention

The present invention relates to isolated polypeptides having phytase activity and isolated nucleic acid sequences encoding the polypeptides. The invention also relates to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences as well 10 as methods for producing the polypeptides. The invention further relates to compositions comprising the polypeptides and methods of use thereof.

Description of the Related Art

Phytases (myo-inositol hexakisphosphate phosphohydrolases, EC 3.1.3.8) catalyze the 15 hydrolysis of phytate (myo-inositol hexakisphosphate) to (1) myo-inositol, (2) mono-, di-, tri-, tetra- and penta-phosphates thereof and (3) inorganic phosphate. Phytase-producing microorganisms include bacteria such as *Bacillus subtilis* (Paver and Jagannathan, 1982, *Journal of Bacteriology* 151: 1102-1108) and *Pseudomonas* (Cosgrove, 1970, Australian Journal of Biological Sciences 23: 1207-1220); yeast such as *Saccharomyces cerevisiae* 20 (Navini and Marcakis, 1984, *Lebensmittel Wissenschaft und Technologie* 17: 24-26; and fungi of the *Aspergillus* genus such as *Aspergillus terreus* (Yamada *et al.*, 1986, *Agricultural Biological Chemistry* 322: 1275-1282).

The cloning and expression of the phytase genes from *Aspergillus niger* var. *awamori* by Piddington *et al.* (1993, *Gene* 133: 55-62) and *Aspergillus niger* (*ficum*) by van 25 Hartingsveldt *et al.* (1993, *Gene* 127: 87-94; EP 420 358) have been disclosed.

Phytic acid is the primary storage form of phosphate in cereal grains, legumes, and oilseeds, such as soy, which are the principal components of animal feeds. However, the presence of phytic acid in animal feeds for monogastric animals is undesirable because the phosphate moieties of phytic acid chelate essential minerals and possibly proteins making them nutritionally unavailable. Furthermore, phytate phosphorus passes through the gastrointestinal tract of monogastric animals and is not metabolized. Since phosphorus is an 30 essential element for the growth of all organisms, livestock feed must be supplemented with

inorganic phosphate. Thus, the art has described the use of phytases in feeds of monogastric animals.

Furthermore, since phytic acid is not metabolized by monogastric animals, it is excreted in manure. The amount of manure produced worldwide has increased significantly resulting from increased livestock production. The disposal of manure has caused an environmental problem in various locations around the world due to the accumulation of phosphate particularly in water. Thus, the art has also described the use of phytases for reducing the amount of phytate in manure.

There is a need in the art for new phytases with improved properties which can be produced in commercially significant quantities.

It is an object of the present invention to provide new phytases with superior properties which can be produced in commercially useful quantities.

Summary of the Invention

The present invention relates to isolated polypeptides having phytase activity selected from the group consisting of:

(a) a polypeptide with an amino acid sequence which has at least 60% identity with the amino acid sequence set forth in SEQ ID NO:2;

(b) a polypeptide which is encoded by a nucleic acid sequence which is capable of hybridizing under medium stringency conditions with (i) the nucleic acid sequence set forth in SEQ ID NO:1 or (ii) its complementary strand;

(c) an allelic form of (a) or (b); and

(d) a fragment of (a), (b), or (c).

The present invention also relates to isolated nucleic acid sequences encoding the polypeptides and to nucleic acid constructs, vectors, and host cells comprising the nucleic acid sequences as well as methods for producing the polypeptides. The present invention further relates to composite feeds and methods of reducing phytate levels.

Brief Description of the Figures

Figure 1 shows an autoradiogram from Southern hybridization analysis of *Thermomyces lanuginosus* CBS 586.94 genomic DNA with a phytase gene probe.

Figure 2 shows the genomic DNA sequence and deduced amino acid sequence of *Thermomyces lanuginosus* CBS 586.94 phytase (SEQ ID NO:1 and SEQ ID NO:2, respectively).

5 Figure 3 shows the alignment of the amino acid sequences for the phytases from *Thermomyces lanuginosus* CBS 586.94 and *Aspergillus niger (ficium)* NRRL 3135 (SEQ ID NO:3).

Figure 4 shows a restriction map of pDM181.

Figure 5 shows a restriction map of pMWR48.

10 Figure 6 shows a comparison of the thermostability of the *Thermomyces lanuginosus* CBS 586.94 and *Aspergillus niger (ficium)* NRRL 3135 phytases.

Figure 7 shows pH-activity profiles of the *Thermomyces lanuginosus* CBS 586.94 and *Aspergillus niger (ficium)* NRRL 3135 phytases.

15 Figure 8 shows temperature-activity profiles of the *Thermomyces lanuginosus* CBS 586.94 and *Aspergillus niger (ficium)* NRRL 3135 phytases.

15

Detailed Description of the Invention

Polypeptides Having Phytase Activity

In a first embodiment, the present invention relates to polypeptides which have an amino acid sequence which has a degree of identity to the amino acid sequence set forth in SEQ ID NO:2 of at least about 60%, preferably at least about 70%, more preferably at least about 80%, even more preferably at least about 90%, most preferably at least 95%, and even most preferably at least about 97%, which qualitatively retain the phytase activity of the polypeptides (hereinafter "homologous polypeptides"). In a preferred embodiment, the homologous polypeptides have an amino acid sequence which differs by five amino acids, preferably by four amino acids, more preferably by three amino acids, even more preferably by two amino acids, and most preferably by one amino acid from the amino acid sequence set forth in SEQ ID NO:2. For purposes of the present invention, the degree of identity between two amino acid sequences is determined by the Clustal method (Higgins, 1989, CABIOS 5: 151-153) with an identity table, a gap penalty of 10, and a gap length penalty of 10.

The amino acid sequences of the homologous polypeptides differ from the amino acid sequence set forth in SEQ ID NO:2 by an insertion or deletion of one or more amino acid residues and/or the substitution of one or more amino acid residues by different amino acid residues. Preferably, amino acid changes are of a minor nature, that is conservative amino acid substitutions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to about 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

Examples of conservative substitutions are within the group of basic amino acids (such as arginine, lysine and histidine), acidic amino acids (such as glutamic acid and aspartic acid), polar amino acids (such as glutamine and asparagine), hydrophobic amino acids (such as leucine, isoleucine and valine), aromatic amino acids (such as phenylalanine, tryptophan and tyrosine), and small amino acids (such as glycine, alanine, serine, threonine and methionine). Amino acid substitutions which do not generally alter the specific activity are known in the art and are described, e.g., by H. Neurath and R.L. Hill, 1979, *In, The Proteins*, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly as well as these in reverse.

In a preferred embodiment, the present invention relates to isolated polypeptides having phytase activity with the amino acid sequence set forth in SEQ ID NO:2, and allelic forms and fragments thereof which retain phytase activity. Preferably, a fragment contains at least 400 amino acid residues, more preferably at least 425 amino acid residues, and most preferably at least 475 amino acid residues.

In a second embodiment, the present invention relates to isolated polypeptides having phytase activity which are encoded by nucleic acid sequences capable of hybridizing under high, medium, or low stringency conditions with an oligonucleotide probe which hybridizes under the same conditions with the nucleic acid sequence set forth in SEQ ID NO:1 or its complementary strand (J. Sambrook, E.F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York), and allelic forms and fragments thereof. Hybridization indicates that the analogous nucleic acid

sequence hybridizes to the oligonucleotide probe corresponding to the polypeptide encoding part of the nucleic acid sequence shown in SEQ ID NO:1, under low to high stringency conditions (for example, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 mg/ml sheared and denatured salmon sperm DNA, and either 50, 35 or 25% formamide for high, medium and low stringencies, respectively), following standard Southern blotting procedures.

The amino acid sequence set forth in SEQ ID NO:2 or a partial amino acid sequence thereof may be used to design an oligonucleotide probe, or a nucleic acid sequence encoding a polypeptide of the present invention, such as the nucleic acid sequence set forth in SEQ ID NO:1, or a subsequence thereof, may be used to identify and clone DNA encoding polypeptides having phytase activity from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic or cDNA of the genus or species of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 15, preferably at least 25, and more preferably at least 40 nucleotides in length. Longer probes can also be used. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with ^{32}P , ^3H , ^{35}S , biotin, or avidin).

Thus, a genomic, cDNA or combinatorial chemical library prepared from such other organisms may be screened for DNA which hybridizes with the probes described above and which encodes a polypeptide having phytase activity. Genomic or other DNA from such other organisms may be separated by agarose or polyacrylamide gel electrophoresis, or other separation techniques. DNA from the libraries or the separated DNA may be transferred to and immobilized on nitrocellulose or other suitable carrier material. In order to identify a clone or DNA which is homologous with SEQ ID NO:1, the carrier material is used in a Southern blot in which the carrier material is finally washed three times for 30 minutes each using 2XSSC, 0.2% SDS at preferably not higher than 50°C, more preferably not higher than 55°C, even more preferably not higher than 60°C, and most preferably not higher than 65°C. Molecules to which the oligonucleotide probe hybridizes under these conditions are detected using X-ray film.

In a preferred embodiment, the isolated polypeptides of the present invention are encoded by nucleic acid sequences capable of hybridizing under medium stringency

conditions with an oligonucleotide probe which hybridizes under the same conditions with the nucleic acid sequence set forth in SEQ ID NO:1 or its complementary strand, and allelic forms and fragments thereof. In a more preferred embodiment, the isolated polypeptides of the present invention are encoded by nucleic acid sequences capable of hybridizing under high stringency conditions with an oligonucleotide probe which hybridizes under the same conditions with the nucleic acid sequence set forth in SEQ ID NO:1 or its complementary strand, and allelic forms and fragments thereof.

The present invention also relates to polypeptides having immunochemical identity or partial immunochemical identity to the polypeptide native to *Thermomyces lanuginosus* CBS 586.94. In this embodiment, a polypeptide of the present invention is used to produce antibodies which are immunoreactive or bind to epitopes of the polypeptide. A polypeptide having immunochemical identity to the polypeptide native to *Thermomyces lanuginosus* CBS 586.94 means that an antiserum containing antibodies against the polypeptide native to *Thermomyces lanuginosus* CBS 586.94 reacts with the other polypeptide in an identical fashion such as total fusion of precipitates, identical precipitate morphology, and/or identical electrophoretic mobility using a specific immunochemical technique. A further explanation of immunochemical identity is described by Axelsen, Bock, and Krøll, *In N.H. Axelsen, J. Krøll, and B. Weeks, editors, A Manual of Quantitative Immunolectrophoresis*, Blackwell Scientific Publications, 1973, Chapter 10. Partial immunochemical identity means that an antiserum containing antibodies against the polypeptide native to *Thermomyces lanuginosus* CBS 586.94 reacts with the other polypeptide in a partially identical fashion such as partial fusion of precipitates, partially identical precipitate morphology, and/or partially identical electrophoretic mobility using a specific immunochemical technique. A further explanation of partial immunochemical identity is described by Bock and Axelsen, *In N.H. Axelsen, J. Krøll, and B. Weeks, editors, A Manual of Quantitative Immunolectrophoresis*, Blackwell Scientific Publications, 1973, Chapter 11. The immunochemical properties are determined by immunological cross-reaction identity tests by the well-known Ouchterlony double immunodiffusion procedure. Specifically, an antiserum against the polypeptide of the invention is raised by immunizing rabbits (or other rodents) according to the procedure described by Harboe and Ingild, *In N.H. Axelsen, J. Krøll, and B. Weeks, editors, A Manual of Quantitative Immunolectrophoresis*, Blackwell Scientific Publications, 1973,

Chapter 23, or Johnstone and Thorpe, *Immunochemistry in Practice*, Blackwell Scientific Publications, 1982 (more specifically pages 27-31).

Polypeptides which are encoded by nucleic acid sequences which are capable of hybridizing with an oligonucleotide probe which hybridizes with the nucleic acid sequence set forth in SEQ ID NO:1 or its complementary strand and allelic forms and fragments thereof, the homologous polypeptides and polypeptides having identical or partially identical immunological properties may be obtained from microorganisms of any genus, preferably from a bacterial or fungal source. Sources for such polypeptides are strains of the genus *Thermomyces* and species thereof available in public depositories. Furthermore, such polypeptides may be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) using the above-mentioned probes. Techniques for isolating microorganisms from natural habitats are well known in the art. The nucleic acid sequence may then be derived by similarly screening a cDNA library of another microorganism, in particular a fungus, such as a strain of an *Aspergillus* sp., in particular a strain of *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* or *Aspergillus oryzae*, a strain of *Trichoderma* sp., in particular a strain of *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* or *Trichoderma viride*, or a strain of a *Fusarium* sp., in particular a strain of *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium graminearum*, *Fusarium oxysporum*, *Fusarium sambucinum*, *Fusarium sulphureum*, or *Fusarium venenatum*, or a strain of a *Humicola* sp., or a strain of an *Aureobasidium* sp., a *Cryptococcus* sp., a *Filibasidium* sp., a *Magnaporthe* sp., a *Myceliophthora* sp., a *Neocallimastix* sp., a *Paecilomyces* sp., a *Piromyces* sp., a *Talaromyces* sp., a *Thermoascus* sp., a *Thielavia* sp., or a *Schizophyllum* sp. Once a nucleic acid sequence encoding a polypeptide has been detected with the probe(s), the sequence may be isolated or cloned by utilizing techniques which are known to those of ordinary skill in the art (see, e.g., Sambrook *et al.*, 1989, *supra*).

The polypeptides of the present invention are preferably obtained from species of *Thermomyces* including, but not limited to, *Thermomyces ibadanensis*, *Thermomyces lanuginosus*, *Thermomyces stellatus*, and *Thermomyces verrucosus*. Strains of these species are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und

Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

In a more preferred embodiment, a polypeptide of the present invention is obtained from *Thermomyces lanuginosus*, and most preferably from *Thermomyces lanuginosus* CBS 586.94 or a mutant strain thereof, e.g., the polypeptide with the amino acid sequence set forth in SEQ ID NO:2.

A polypeptide of the present invention may further be obtained from other fungi which are synonyms of *Thermomyces* as described by S.C. Jong, J.M. Birmingham, and G. Ma in *ATCC Names of Industrial Fungi*, American Type Culture Collection, Rockville, Maryland, 1994 or M.B. Ellis in *Dematiaceous Hyphomycetes*, Commonwealth Mycological Institute, Surrey, England, 1971. For example, synonyms of *Thermomyces lanuginosus* include *Acremoniella thermophila*, *Humicola lanuginosa*, *Monotospora lanuginosa*, and *Sepedonium lanuginosum*. The present invention also encompasses phytases obtained from fungi which are teleomorphs of *Thermomyces*. The genus *Thermomyces* is a terrestrial member of the group of dematiaceous hyphomycete fungi. Colonies are effuse, cottony or velvety, and grey, greenish grey, buff, dark blackish brown, or black. Mycelia are partly superficial, partly immersed. Conidiophores are micronematous or semi-macronematous, mononematous, unbranched or irregularly branched, straight or flexuous, colorless or brown, and smooth. Conidiogenous cells are monoblastic, integrated and terminal or discrete, determinate, cylindrical or lageniform. Conidia are solitary, dry, acrogenous, simple, spherical to subspherical or angular and lobed, pale to dark blackish brown, smooth or verrucose, and O-septate. No phialidic state is known.

For purposes of the present invention, the term "obtained from" as used herein in connection with a given source shall mean that the polypeptide is produced by the source or by a cell in which a gene from the source has been inserted.

As defined herein, an "isolated" polypeptide is a polypeptide which is essentially free of other non-phytase polypeptides, e.g., at least about 20% pure, preferably at least about 40% pure, more preferably about 60% pure, even more preferably about 80% pure, most preferably about 90% pure, and even most preferably about 95% pure, as determined by SDS-PAGE.

The polypeptides of the present invention are characterized as having a high activity at high temperatures. More specifically, these polypeptides have maximum phytase activity

near 65°C and partial activity even at 75°C. In contrast, the *Aspergillus niger* phytase is essentially inactive at 65°C.

Nucleic Acid Sequences

5 The present invention also relates to isolated nucleic acid sequences which encode a polypeptide of the present invention. In a preferred embodiment, the nucleic acid sequence encodes a polypeptide obtained from *Thermomyces*, e.g., *Thermomyces lanuginosus*, and in a more preferred embodiment, the nucleic acid sequence is obtained from *Thermomyces lanuginosus* CBS 586:94, e.g., the nucleic acid sequence set forth in SEQ ID NO:1. In a
10 more preferred embodiment, the nucleic acid sequence is the sequence contained in plasmid pMWR46 which is contained in *Escherichia coli* NRRL B-21527. The present invention also encompasses nucleic acid sequences which encode a polypeptide having the amino acid sequence set forth in SEQ ID NO:2, which differ from SEQ ID NO:1 by virtue of the degeneracy of the genetic code. The present invention also relates to subsequences of SEQ
15 ID NO:1 which encode a fragment of SEQ ID NO:2 which retains phytase activity. Preferably, a subsequence contains at least 1200 nucleotides, more preferably at least 1275 nucleotides, and most preferably at least 1425 nucleotides.

20 As described above, the nucleic acid sequences may be obtained from microorganisms which are synonyms or teleomorphs of *Thermomyces* as defined by M.B. Ellis, 1971, *supra* or Jong *et al.*, 1994, *supra*.

25 The techniques used to isolate or clone a nucleic acid sequence encoding a polypeptide are known in the art and include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the nucleic acid sequences of the present invention from such genomic DNA can be effected, e.g., by using the well known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleic acid sequence-based amplification (NASBA) may be used. The nucleic acid sequence may be cloned from a strain of the *Thermomyces* producing the polypeptide, or another or related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the nucleic acid sequence.

The term "isolated" nucleic acid sequence as used herein refers to a nucleic acid sequence which is essentially free of other nucleic acid sequences, *e.g.*, at least about 20% pure, preferably at least about 40% pure, more preferably about 60% pure, even more preferably about 80% pure, most preferably about 90% pure, and even most preferably about 95% pure, as determined by agarose gel electrophoresis. For example, an isolated nucleic acid sequence can be obtained by standard cloning procedures used in genetic engineering to relocate the nucleic acid sequence from its natural location to a different site where it will be reproduced. The cloning procedures may involve excision and isolation of a desired nucleic acid fragment comprising the nucleic acid sequence encoding the polypeptide, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into a host cell where multiple copies or clones of the nucleic acid sequence will be replicated. The nucleic acid sequence may be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

The present invention also relates to nucleic acid sequences which have a nucleic acid sequence which has a degree of identity to the nucleic acid sequence set forth in SEQ ID NO:1 of at least about 60%, preferably at least about 70%, more preferably at least about 80%, even more preferably at least about 90%, most preferably at least about 95%, and even most preferably at least about 97%, which encode an active polypeptide. For purposes of the present invention, the degree of identity between two nucleic acid sequences is determined by the Clustal method (Higgins, 1989, CABIOS 5: 151-153) with an identity table, a gap penalty of 10, and a gap length penalty of 10.

Modification of the nucleic acid sequence encoding the polypeptide may be necessary for the synthesis of polypeptides substantially similar to the polypeptide. The term "substantially similar" to the polypeptide refers to non-naturally occurring forms of the polypeptide. These polypeptides may differ in some engineered way from the polypeptide isolated from its native source. For example, it may be of interest to synthesize variants of the polypeptide where the variants differ in specific activity, thermostability, pH optimum, or the like using, *e.g.*, site-directed mutagenesis. The analogous sequence may be constructed on the basis of the nucleic acid sequence presented as the polypeptide encoding part of SEQ ID NO:1, *e.g.*, a sub-sequence thereof, and/or by introduction of nucleotide substitutions which do not give rise to another amino acid sequence of the polypeptide encoded by the nucleic acid sequence, but which corresponds to the codon usage of the host

organism intended for production of the enzyme, or by introduction of nucleotide substitutions which may give rise to a different amino acid sequence. For a general description of nucleotide substitution, see, e.g., Ford *et al.*, 1991, *Protein Expression and Purification* 2:95-107.

5 It will be apparent to those skilled in the art that such substitutions can be made outside the regions critical to the function of the molecule and still result in an active polypeptide. Amino acid residues essential to the activity of the polypeptide encoded by the isolated nucleic acid sequence of the invention, and therefore preferably not subject to substitution, may be identified according to procedures known in the art, such as site-directed
10 mutagenesis or alanine-scanning mutagenesis (see, e.g., Cunningham and Wells, 1989, *Science* 244:1081-1085). In the latter technique mutations are introduced at every positively charged residue in the molecule, and the resultant mutant molecules are tested for phytase activity to identify amino acid residues that are critical to the activity of the molecule. Sites
15 of substrate-enzyme interaction can also be determined by analysis of three-dimensional structure as determined by such techniques as nuclear magnetic resonance analysis, crystallography or photoaffinity labelling (see, e.g., de Vos *et al.*, 1992, *Science* 255, 306-312; Smith *et al.*, 1992, *Journal of Molecular Biology* 224:899-904; Wlodaver *et al.*, 1992, *FEBS Letters* 309, 59-64).

20 Polypeptides of the present invention also include fused polypeptides or cleavable fusion polypeptides in which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide or fragment thereof. A fused polypeptide is produced by fusing a nucleic acid sequence (or a portion thereof) encoding another polypeptide to a nucleic acid sequence (or a portion thereof) of the present invention. Techniques for producing fusion polypeptides are known in the art, and include, ligating the coding sequences encoding the polypeptides
25 so that they are in frame and that expression of the fused polypeptide is under control of the same promoter(s) and terminator.

30 The present invention also relates to isolated polypeptides having phytase activity encoded by nucleic acid sequences which are capable of hybridizing under medium stringency conditions with an oligonucleotide probe which hybridizes under the same conditions with the nucleic acid sequence set forth in SEQ ID NO:1 or its complementary strand (Sambrook *et al.*, 1989, *supra*). Hybridization indicates that the analogous nucleic acid sequence

hybridizes to the oligonucleotide probe corresponding to the polypeptide encoding part of the nucleic acid sequence shown in SEQ ID NO:1 under standard conditions.

As disclosed earlier, the amino acid sequence set forth in SEQ ID NO:2 or a partial amino acid sequence thereof may be used to design an oligonucleotide probe, or a nucleic acid sequence encoding a polypeptide of the present invention, such as the nucleic acid sequence set forth in SEQ ID NO:1, or a subsequence thereof can also be used as a probe, to isolate homologous genes of any genus or species.

Nucleic Acid Constructs

The present invention also relates to nucleic acid constructs comprising a nucleic acid sequence of the present invention operably linked to one or more control sequences capable of directing the expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences.

"Nucleic acid construct" is defined herein as a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid which are combined and juxtaposed in a manner which would not otherwise exist in nature. The term nucleic acid construct may be synonymous with the term expression cassette when the nucleic acid construct contains all the control sequences required for expression of a coding sequence of the present invention. The term "coding sequence" as defined herein is a sequence which is transcribed into mRNA and translated into a polypeptide of the present invention when placed under the control of the above mentioned control sequences. The boundaries of the coding sequence are generally determined by a translation start codon ATG at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to, DNA, cDNA, and recombinant nucleic acid sequences.

An isolated nucleic acid sequence encoding a polypeptide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the nucleic acid sequence encoding a polypeptide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying nucleic acid sequences utilizing cloning methods are well known in the art.

The term "control sequences" is defined herein to include all components which are necessary or advantageous for expression of the coding sequence of the nucleic acid

sequence. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, a polyadenylation sequence, a propeptide sequence, a promoter, a signal sequence, and a transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

The control sequence may be an appropriate promoter sequence, a nucleic acid sequence which is recognized by a host cell for expression of the nucleic acid sequence. The promoter sequence contains transcriptional control sequences which mediate the expression of the polypeptide. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention, especially in a bacterial host cell, are the promoters obtained from the *E. coli lac* operon, the *Streptomyces coelicolor* agarase gene (*dagA*), the *Bacillus subtilis* levansucrase gene (*sacB*), the *Bacillus licheniformis* alpha-amylase gene (*amyL*), the *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), the *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), the *Bacillus licheniformis* penicillinase gene (*penP*), the *Bacillus subtilis* *xylA* and *xylB* genes, and the prokaryotic beta-lactamase gene (Villa-Kamaroff *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75:3727-3731), as well as the *tac* promoter (DeBoer *et al.*, 1983, *Proceedings of the National Academy of Sciences USA* 80:21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242:74-94; and in Sambrook *et al.*, 1989, *supra*.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei*

lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Fusarium oxysporum* trypsin-like protease (as described in U.S. Patent No. 4,288,627, which is incorporated herein by reference), and mutant, truncated, and hybrid promoters thereof. Particularly preferred promoters for use in filamentous fungal host cells are the TAKA amylase, NA2-tpi (a hybrid of the promoters from the genes encoding *Aspergillus niger* neutral α -amylase and *Aspergillus oryzae* triose phosphate isomerase), and *glaA* promoters.

In a yeast host, useful promoters are obtained from the *Saccharomyces cerevisiae* enolase (ENO-1) gene, the *Saccharomyces cerevisiae* galactokinase gene (GAL1), the *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase genes (ADH2/GAP), and the *Saccharomyces cerevisiae* 3-phosphoglycerate kinase gene. Other useful promoters for yeast host cells are described by Romanos *et al.*, 1992, *Yeast* 8:423-488. In a mammalian host cell, useful promoters include viral promoters such as those from Simian Virus 40 (SV40), Rous sarcoma virus (RSV), adenovirus, and bovine papilloma virus (BPV).

The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice may be used in the present invention.

Preferred terminators for filamentous fungal host cells are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease.

Preferred terminators for yeast host cells are obtained from the genes encoding *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), or *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos *et al.*, 1992, *supra*. Terminator sequences are well known in the art for mammalian host cells.

The control sequence may also be a suitable leader sequence, a nontranslated region of a mRNA which is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide.

Any leader sequence which is functional in the host cell of choice may be used in the present invention.

Preferred leaders for filamentous fungal host cells are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase and *Aspergillus oryzae* triose phosphate isomerase.

Suitable leaders for yeast host cells are obtained from the *Saccharomyces cerevisiae* enolase (ENO-1) gene, the *Saccharomyces cerevisiae* 3-phosphoglycerate kinase gene, the *Saccharomyces cerevisiae* alpha-factor, and the *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase genes (ADH2/GAP).

The control sequence may also be a polyadenylation sequence, a sequence which is operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention.

Preferred polyadenylation sequences for filamentous fungal host cells are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, and *Aspergillus niger* alpha-glucosidase.

Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, *Molecular Cellular Biology* 15:5983-5990. Polyadenylation sequences are well known in the art for mammalian host cells.

The control sequence may also be a signal peptide coding region, which codes for an amino acid sequence linked to the amino terminus of the polypeptide which can direct the expressed polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to that portion of the coding sequence which encodes the secreted polypeptide. The foreign signal peptide coding region may be required where the coding sequence does not normally contain a signal peptide coding region. Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to obtain enhanced secretion of the phytase relative to the natural signal peptide coding region normally associated with the coding sequence. The signal peptide

coding region may be obtained from a glucoamylase or an amylase gene from an *Aspergillus* species, a lipase or proteinase gene from a *Rhizomucor* species, the gene for the alpha-factor from *Saccharomyces cerevisiae*, an amylase or a protease gene from a *Bacillus* species, or the calf prochymosin gene. However, any signal peptide coding region capable of directing the expressed phytase into the secretory pathway of a host cell of choice may be used in the present invention.

An effective signal peptide coding region for bacterial host cells is the signal peptide coding region obtained from the maltogenic amylase gene from *Bacillus* NCIB 11837, the *Bacillus stearothermophilus* alpha-amylase gene, the *Bacillus licheniformis* subtilisin gene, the *Bacillus licheniformis* beta-lactamase gene, the *Bacillus stearothermophilus* neutral proteases genes (*nprT*, *nprS*, *nprM*), and the *Bacillus subtilis* PrsA gene. Further signal peptides are described by Simonen and Palva, 1993, *Microbiological Reviews* 57:109-137.

An effective signal peptide coding region for filamentous fungal host cells is the signal peptide coding region obtained from *Aspergillus oryzae* TAKA amylase gene, *Aspergillus niger* neutral amylase gene, the *Rhizomucor miehei* aspartic proteinase gene, the *Humicola lanuginosa* cellulase gene, or the *Rhizomucor miehei* lipase gene.

Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* a-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding regions are described by Romanos *et al.*, 1992, *supra*.

The control sequence may also be a propeptide coding region, which codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the *Bacillus subtilis* alkaline protease gene (*aprE*), the *Bacillus subtilis* neutral protease gene (*nprT*), the *Saccharomyces cerevisiae* alpha-factor gene, or the *Myceliophthora thermophila* laccase gene (WO 95/33836).

The nucleic acid constructs of the present invention may also comprise one or more nucleic acid sequences which encode one or more factors that are advantageous in the expression of the polypeptide, *e.g.*, an activator (*e.g.*, a trans-acting factor), a chaperone, and a processing protease. Any factor that is functional in the host cell of choice may be

used in the present invention. The nucleic acids encoding one or more of these factors are not necessarily in tandem with the nucleic acid sequence encoding the polypeptide.

An activator is a protein which activates transcription of a nucleic acid sequence encoding a polypeptide (Kudla *et al.*, 1990, *EMBO Journal* 9: 1355-1364; Jarai and Buxton, 5 1994, *Current Genetics* 26: 2238-244; Verdier, 1990, *Yeast* 6: 271-297). The nucleic acid sequence encoding an activator may be obtained from the genes encoding *Bacillus stearothermophilus* NprA (*nprA*), *Saccharomyces cerevisiae* heme activator protein 1 (*hap1*), *Saccharomyces cerevisiae* galactose metabolizing protein 4 (*gal4*), and *Aspergillus nidulans* ammonia regulation protein (*areA*). For further examples, see Verdier, 1990, *supra* and 10 MacKenzie *et al.*, 1993, *Journal of General Microbiology* 139: 2295-2307.

A chaperone is a protein which assists another polypeptide in folding properly (Hartl *et al.*, 1994, *TIBS* 19: 20-25; Bergeron *et al.*, 1994, *TIBS* 19: 124-128; Demolder *et al.*, 1994, *Journal of Biotechnology* 32: 179-189; Craig, 1993, *Science* 260: 1902-1903; Gething 15 and Sambrook, 1992, *Nature* 355: 33-45; Puig and Gilbert, 1994, *Journal of Biological Chemistry* 269: 7764-7771; Wang and Tsou, 1993, *The FASEB Journal* 7: 1515-11157; Robinson *et al.*, 1994, *Bio/Technology* 1: 381-384). The nucleic acid sequence encoding a chaperone may be obtained from the genes encoding *Bacillus subtilis* GroE proteins, *Aspergillus oryzae* protein disulphide isomerase, *Saccharomyces cerevisiae* calnexin, *Saccharomyces cerevisiae* BiP/GRP78, and *Saccharomyces cerevisiae* Hsp70. For further 20 examples, see Gething and Sambrook, 1992, *supra*, and Hartl *et al.*, 1994, *supra*.

A processing protease is a protease that cleaves a propeptide to generate a mature biochemically active polypeptide (Enderlin and Ogrydziak, 1994, *Yeast* 10: 67-79; Fuller *et al.*, 1989, *Proceedings of the National Academy of Sciences USA* 86: 1434-1438; Julius *et al.*, 1984, *Cell* 37: 1075-1089; Julius *et al.*, 1983, *Cell* 32: 839-852). The nucleic acid 25 sequence encoding a processing protease may be obtained from the genes encoding *Saccharomyces cerevisiae* dipeptidylaminopeptidase, *Saccharomyces cerevisiae* Kex2, and *Yarrowia lipolytica* dibasic processing endoprotease (*xpr6*).

It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory 30 systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory systems in prokaryotic systems would include the *lac*, *tac*, and *trp* operator

systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and the *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences. Other examples of regulatory sequences are those which allow for gene amplification. In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the metallothionein genes which are amplified with heavy metals. In these cases, the nucleic acid sequence encoding the polypeptide would be operably linked with the regulatory sequence.

10 Expression Vectors

The present invention also relates to recombinant expression vectors comprising a nucleic acid sequence of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression, and possibly secretion.

The recombinant expression vector may be any vector (*e.g.*, a plasmid or virus) which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the nucleic acid sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids. The vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. The vector system may be a single vector

or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon.

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. A frequently used mammalian marker is the dihydrofolate reductase gene. Suitable markers for yeast host cells are *ADE2*, *HIS3*, *LEU2*, *LYS2*, *MET3*, *TRP1*, and *URA3*. A selectable marker for use in a filamentous fungal host cell may be selected from the group including, but not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar* (phosphinothricin acetyltransferase), *hygB* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenyltransferase), *trpC* (anthranilate synthase), and glufosinate resistance markers, as well as equivalents from other species. Preferred for use in an *Aspergillus* cell are the *amdS* and *pyrG* genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the *bar* gene of *Streptomyces hygroscopicus*. Furthermore, selection may be accomplished by co-transformation, e.g., as described in WO 91/17243, where the selectable marker is on a separate vector.

The vectors of the present invention preferably contain an element(s) that permits stable integration of the vector into the host cell genome or autonomous replication of the vector in the cell independent of the genome of the cell.

The vectors of the present invention may be integrated into the host cell genome when introduced into a host cell. For integration, the vector may rely on the nucleic acid sequence encoding the polypeptide or any other element of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to

5

1,500 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination. These nucleic acid sequences may be any sequence that is homologous with a target sequence in the genome of the host cell, and, furthermore, may be non-encoding or encoding sequences.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAMβ1 permitting replication in *Bacillus*. Examples of origin of replications for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6. The origin of replication may be one having a mutation which makes its functioning temperature-sensitive in the host cell (see, e.g., Ehrlich, 1978, *Proceedings of the National Academy of Sciences USA* 75:1433).

More than one copy of a nucleic acid sequence encoding a polypeptide of the present invention may be inserted into the host cell to amplify expression of the nucleic acid sequence. Stable amplification of the nucleic acid sequence can be obtained by integrating at least one additional copy of the sequence into the host cell genome using methods well known in the art and selecting for transformants.

The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, e.g., Sambrook *et al.*, 1989, *supra*).

Host Cells

The present invention also relates to recombinant host cells, comprising a nucleic acid sequence of the invention, which are advantageously used in the recombinant production of the polypeptides. The term "host cell" encompasses any progeny of a parent cell which is not identical to the parent cell due to mutations that occur during replication.

The cell is preferably transformed with a vector comprising a nucleic acid sequence of the invention followed by integration of the vector into the host chromosome. "Transformation" means introducing a vector comprising a nucleic acid sequence of the present invention into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector. Integration is generally considered to be an advantage as the nucleic acid sequence is more likely to be stably maintained in the cell. Integration of the vector into the host chromosome may occur by homologous or non-homologous recombination as described above.

The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source. The host cell may be a unicellular microorganism, e.g., a prokaryote, or a non-unicellular microorganism, e.g., a eukaryote. Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a *Bacillus* cell, e.g., *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus laetus*, *Bacillus lentinus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Bacillus thuringiensis*; or a *Streptomyces* cell, e.g., *Streptomyces lividans* or *Streptomyces murinus*, or gram negative bacteria such as *E. coli* and *Pseudomonas* sp. In a preferred embodiment, the bacterial host cell is a *Bacillus lentinus*, *Bacillus licheniformis*, *Bacillus stearothermophilus* or *Bacillus subtilis* cell. The transformation of a bacterial host cell may, for instance, be effected by protoplast transformation (see, e.g., Chang and Cohen, 1979, *Molecular General Genetics* 168:111-115), by using competent cells (see, e.g., Young and Spizizin, 1961, *Journal of Bacteriology* 81:823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56:209-221), by electroporation (see, e.g., Shigekawa and Dower, 1988, *Biotechniques* 6:742-751), or by conjugation (see, e.g., Koehler and Thorne, 1987, *Journal of Bacteriology* 169:5771-5278).

The host cell may be a eukaryote, such as a mammalian cell, an insect cell, a plant cell or a fungal cell. Useful mammalian cells include Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, COS cells, or any number of other immortalized cell lines available, e.g., from the American Type Culture Collection.

In a preferred embodiment, the host cell is a fungal cell. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth *et al.*, *In, Ainsworth and Bisby's Dictionary of The Fungi*, 8th

edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth *et al.*, 1995, *supra*, page 171) and all mitosporic fungi (Hawksworth *et al.*, 1995, *supra*). Representative groups of Ascomycota include, e.g., *Neurospora*, *Eupenicillium* (= *Penicillium*), *Emericella* (= *Aspergillus*), *Eurotium* (= *Aspergillus*), and the true yeasts listed below. Examples of Basidiomycota include mushrooms, rusts, and smuts. Representative groups of Chytridiomycota include, e.g., *Allomyces*, *Blastocladiella*, *Coelomomyces*, and aquatic fungi. Representative groups of Oomycota include, e.g., Saprolegniomycetous aquatic fungi (water molds) such as *Achlya*. Examples of mitosporic fungi include *Aspergillus*, *Penicillium*, *Candida*, and *Alternaria*. Representative groups of Zygomycota include, e.g., *Rhizopus* and *Mucor*.

In a preferred embodiment, the fungal host cell is a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). The ascosporogenous yeasts are divided into the families Spermophthoraceae and Saccharomycetaceae. The latter is comprised of four subfamilies, Schizosaccharomycoideae (e.g., genus *Schizosaccharomyces*), Nadsonioideae, Lipomycoideae, and Saccharomycoideae (e.g., genera *Pichia*, *Kluyveromyces* and *Saccharomyces*). The basidiosporogenous yeasts include the genera *Leucosporidium*, *Rhodosporidium*, *Sporidiobolus*, *Filobasidium*, and *Filobasidiella*. Yeast belonging to the Fungi Imperfecti are divided into two families, Sporobolomycetaceae (e.g., genera *Sorobolomyces* and *Bullera*) and Cryptococcaceae (e.g., genus *Candida*). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980. The biology of yeast and manipulation of yeast genetics are well known in the art (see, e.g., *Biochemistry and Genetics of Yeast*, Bacil, M., Horecker, B.J., and Stopani, A.O.M., editors, 2nd edition, 1987; *The Yeasts*, Rose, A.H., and Harrison, J.S., editors, 2nd edition, 1987; and *The Molecular Biology of the Yeast Saccharomyces*, Strathern *et al.*, editors, 1981).

In a more preferred embodiment, the yeast host cell is a cell of a species of *Candida*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, or *Yarrowia*.

In a most preferred embodiment, the yeast host cell is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces*

douglasii, *Saccharomyces kluyveri*, *Saccharomyces norbensis* or *Saccharomyces oviformis* cell. In another most preferred embodiment, the yeast host cell is a *Kluyveromyces lactis* cell. In another most preferred embodiment, the yeast host cell is a *Yarrowia lipolytica* cell.

In a preferred embodiment, the fungal host cell is a filamentous fungal cell. 5 "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota (as defined by Hawksworth *et al.*, 1995, *supra*). The filamentous fungi are characterized by a mycelial wall composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides. Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In contrast, vegetative growth by yeasts such as 10 *Saccharomyces cerevisiae* is by budding of a unicellular thallus and carbon catabolism may be fermentative. In a more preferred embodiment, the filamentous fungal host cell is a cell of a species of, but not limited to, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, and *Trichoderma*.

In an even more preferred embodiment, the filamentous fungal host cell is an 15 *Aspergillus* cell. In another even more preferred embodiment, the filamentous fungal host cell is an *Acremonium* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Fusarium* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Humicola* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Mucor* cell. In another even more preferred 20 embodiment, the filamentous fungal host cell is a *Myceliophthora* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Neurospora* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Penicillium* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Thielavia* cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Tolypocladium* 25 cell. In another even more preferred embodiment, the filamentous fungal host cell is a *Trichoderma* cell.

In a most preferred embodiment, the filamentous fungal host cell is an *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus niger* or *Aspergillus oryzae* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium graminearum*, *Fusarium oxysporum*, *Fusarium sambucinum*, *Fusarium sulphureum*, or *Fusarium venenatum* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Humicola insolens* or *Humicola lanuginosa*

cell. In another most preferred embodiment, the filamentous fungal host cell is a *Mucor miehei* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Myceliophthora thermophilum* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Neurospora crassa* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Penicillium purpurogenum* cell. In another most preferred embodiment, the filamentous fungal host cell is a *Thielavia terrestris* cell. In another most preferred embodiment, the *Trichoderma* cell is a *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* or *Trichoderma viride* cell.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and Yelton *et al.*, 1984, *Proceedings of the National Academy of Sciences USA* 81:1470-1474. A suitable method of transforming *Fusarium* species is described by Malardier *et al.*, 1989, *Gene* 78:147-156 or in copending US Serial No. 08/269,449. Yeast may be transformed using the procedures described by Becker and Guarente, *In Abelson, J.N. and Simon, M.I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153:163; and Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75:1920. Mammalian cells may be transformed by direct uptake using the calcium phosphate precipitation method of Graham and Van der Eb (1978, *Virology* 52:546).

Methods of Production

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a *Thermomyces* strain to produce a supernatant comprising the polypeptide; and (b) recovering the polypeptide.

The present invention also relates to methods for producing a polypeptide of the present invention comprising (a) cultivating a host cell under conditions conducive to expression of the polypeptide; and (b) recovering the polypeptide.

In both methods, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous,

batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art (see, e.g., references for bacteria and yeast; Bennett, J.W. and LaSure, L., editors, *More Gene Manipulations in Fungi*, Academic Press, CA, 1991). Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g., in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it is recovered from cell lysates.

The polypeptides may be detected using methods known in the art that are specific for the polypeptides. These detection methods may include use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate. For example, an enzyme assay may be used to determine the activity of the polypeptide. Procedures for determining phytase activity are known in the art and include, e.g., the assay of inorganic phosphate liberated from phytic acid.

The resulting polypeptide may be recovered by methods known in the art. For example, the polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation. The recovered polypeptide may then be further purified by a variety of chromatographic procedures, e.g., ion exchange chromatography, gel filtration chromatography, affinity chromatography, or the like.

The polypeptides of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing (IEF), differential solubility (e.g., ammonium sulfate precipitation), or extraction (see, e.g., *Protein Purification*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

30 Polypeptide Compositions

In a still further aspect, the present invention relates to polypeptide compositions which are enriched in a polypeptide of the present invention. In the present context, the term

"enriched" is intended to indicate that the phytase activity of the polypeptide composition has been increased, *e.g.*, with an enrichment factor of 1.1, conveniently due to addition of a polypeptide of the invention.

The polypeptide composition may be one which comprises a polypeptide of the invention as the major enzymatic component, *e.g.*, a mono-component polypeptide composition. Alternatively, the composition may comprise multiple enzymatic activities, such as an aminopeptidase, an amylase, a carbohydrolase, a carboxypeptidase, a catalase, a cellulase, a chitinase, a cutinase, a cyclodextrin glycosyltransferase, a deoxyribonuclease, an esterase, an alpha-galactosidase, a beta-galactosidase, a glucoamylase, an alpha-glucosidase, a beta-glucosidase, a haloperoxidase, an invertase, a laccase, a lipase, a mannosidase, an oxidase, a pectinolytic enzyme, a peroxidase, a phytase, a polyphenoloxidase, a proteolytic enzyme, a ribonuclease, or a xylanase. The additional enzyme(s) may be producible by means of a microorganism belonging to the genus *Aspergillus*, preferably *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus niger*, or *Aspergillus oryzae*, or *Trichoderma*, *Humicola*, preferably *Humicola insolens*, or *Fusarium*, preferably *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium graminearum*, *Fusarium oxysporum*, *Fusarium sambucinum*, *Fusarium sulphureum*, or *Fusarium venenatum*.

The polypeptide compositions may be prepared in accordance with methods known in the art and may be in the form of a liquid or a dry composition. The polypeptide to be included in the composition may be stabilized in accordance with methods known in the art.

Examples are given below of preferred uses of the polypeptide compositions of the invention. The dosage of the polypeptide composition of the invention and other conditions under which the composition is used may be determined on the basis of methods known in the art.

25

Uses

The present invention also relates generally to the use of the polypeptide of the present invention for catalyzing the liberation of inorganic phosphate from phytate or phytic acid.

30

More specifically, the polypeptides may be used in human food or animal feed compositions or as additives for such preparations, wherein the phytase improves

digestibility, promotes growth, and improves food and feed utilization or its conversion efficiency.

A "feed composition" and a "food composition," respectively, means any natural or artificial diet, meal or the like or components of such meals intended or suitable for being eaten, taken in, digested, by an animal and a human being, respectively.

A "food or feed additive" is an essentially pure compound or a multi component composition intended for or suitable for being added to food or feed. It usually comprises one or more compounds such as vitamins, minerals or feed enhancing enzymes and suitable carriers and/or excipients, and it is usually provided in a form that is suitable for being added to animal feed.

The invention also relates to feed and food compositions and additives therefore comprising a polypeptide of the invention.

An effective amount of the polypeptide in food or feed is from about 10-20,000; preferably from about 10 to 15,000, more preferably from about 10 to 10,000, in particular from about 100 to 5,000, especially from about 100 to about 2,000 U/kg feed or food.

The invention also relates to a method for reducing phytate levels in animal manure, comprising feeding an animal a feed comprising an effective amount of a polypeptide of the invention.

Also within the scope of the invention is the use of a polypeptide of the invention during the preparation of food or feed preparations or additives, *i.e.*, the polypeptide exerts its phytase activity during the manufacture only and is not active in the final food or feed product. This aspect is relevant for instance in baking.

The polypeptides of the invention can also be used in soy processing and corn steeping. Corn steep liquor is a universally used ingredient in fermentation. The corn steep liquor causes sedimentation because of its phytate content, which results in problems in recovering or purifying the product. The pre-treatment of the steeping corn with a polypeptide of the present invention eliminates these problems.

The present invention also relates to the use of the polypeptides in methods for liquefying a starch, comprising (a) treating the starch with a polypeptide of claim 1 prior to or simultaneously with liquefying; (b) adding an α -amylase to the starch; and (c) reacting the starch of step (b) for a time and at a temperature effective to liquefy the starch. In this process, the polypeptide catalyzes the hydrolysis of phytate associated with the starch.

The present invention is further described by the following examples which should not be construed as limiting the scope of the invention.

Examples

5

Example 1: *Thermomyces lanuginosus* CBS 586.94 Genomic DNA Extraction

Thermomyces lanuginosus CBS 586.94 was grown in 25 ml of 0.5% yeast extract-2% glucose (YEG) medium for 24 hours at 32°C and 250 rpm. Mycelia were then collected by filtration through Miracloth (Calbiochem, La Jolla, CA) and washed once with 25 ml of 10 mM Tris-1 mM EDTA (TE) buffer. Excess buffer was drained from the mycelia which were subsequently frozen in liquid nitrogen. The frozen mycelia were ground to a fine powder in an electric coffee grinder, and the powder was added to 20 ml of TE buffer and 5 ml of 20% w/v sodium dodecylsulfate (SDS) in a disposable plastic centrifuge tube. The mixture was gently inverted several times to ensure mixing, and extracted twice with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v). Sodium acetate (3 M solution) was added to give a final concentration of 0.3 M and the nucleic acids were precipitated with 2.5 volumes of ice cold ethanol. The tube was centrifuged at 15,000 x g for 30 minutes and the pellet was allowed to air dry for 30 minutes before resuspension in 0.5 ml of TE buffer. DNase-free ribonuclease A was added to a concentration of 100 µg/ml and the mixture was incubated at 37°C for 30 min. Proteinase K (200 µg/ml) was then added and the mixture was incubated an additional hour at 37°C. Finally, the mixture was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v) before precipitating the DNA with sodium acetate and ethanol. The DNA pellet was dried under vacuum, resuspended in TE buffer, and stored at 4°C.

25

Example 2: Hybridization Analysis of Genomic DNA

The total cellular DNA sample prepared as described in Example 1 was analyzed by Southern hybridization (Maniatis *et al.*, 1982, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York). Approximately 5 µg of the DNA sample were digested with *Bam*H1 or *Pst*I and fractionated by size on a 1% agarose gel. The gel was photographed under short wavelength UV light and soaked for 15 minutes in 0.5 M NaOH-1.5 M NaCl followed by 15 minutes in 1 M Tris-HCl pH 8-1.5 M NaCl.

DNA in the gel was transferred onto a Nytran™ hybridization membrane (Schleicher & Schuell, Keene, NH) by capillary blotting in 20 X SSPE (3 M sodium chloride-0.2 M sodium dibasic phosphate-0.02 M disodium EDTA) according to Davis et al. (1980, Advanced Bacterial Genetics, A Manual for Genetic Engineering, Cold Spring Harbor Press, Cold Spring Harbor, New York). The membrane was baked for 2 hours at 80°C under vacuum and was soaked for 2 hours in the following hybridization buffer at 45°C with gentle agitation: 5 X SSPE, 25% formamide (v/v), 0.3% SDS, and 200 µg/ml denatured and sheared salmon testes DNA. A phytase-specific probe fragment (approximately 1.6 kb) was radiolabeled by nick translation (Maniatis et al., 1982, *supra*) with $\alpha^{[32]P]$ dCTP (Amersham, Arlington Heights, IL) and added to the hybridization buffer at an activity of approximately 1×10^6 cpm per ml of buffer. The mixture was incubated with the membrane overnight at 45°C in a shaking water bath. Following incubation, the membrane was washed once in 1.0 X SSPE with 0.1% SDS at 45°C followed by two washes in 1.0 X SSPE (no SDS) at the same temperature. The membrane was dried on a paper towel for 15 minutes, then wrapped in Saran-Wrap™ and exposed to X-ray film overnight at -70°C with intensifying screens (Kodak, Rochester, NY).

Southern blotting indicates that the probe can be used as a probe to identify and clone the phytase gene from *Thermomyces lanuginosus* CBS 586.94 as shown in Figure 1.

20 Example 3: DNA Libraries and Identification of Phytase Clones

Genomic DNA libraries were constructed using the bacteriophage cloning vector λZipLox (Life Technologies, Gaithersburg, MD) with *E. coli* Y1090ZL cells (Life Technologies, Gaithersburg, MD) as a host for plating and purification of recombinant bacteriophage and *E. coli* DH10Bzip (Life Technologies, Gaithersburg, MD) for excision of individual pZL1-phytase clones. Total cellular DNA was partially digested with *Tsp*509I and size-fractionated on 1% agarose gels. DNA fragments migrating in the size range 3-7 kb were excised and eluted from the gel using Prep-a-Gene reagents (BioRad Laboratories, Hercules, CA). The eluted DNA fragments were ligated with *Eco*RI-cleaved and dephosphorylated λZipLox vector arms (Life Technologies, Gaithersburg, MD), and the ligation mixtures were packaged using commercial packaging extracts (Stratagene, La Jolla, CA). The packaged DNA libraries were plated and amplified in *E. coli* Y1090ZL cells (Life Technologies, Gaithersburg, MD). Approximately 30,000 plaques from the library were

screened by plaque-hybridization with the radiolabeled phytase probe described in Example 2. One positive clone which hybridizes strongly to the probe was picked and purified twice in *E. coli* Y1090ZL cells. The phytase clone was subsequently excised from the λZipLox vector as pZL1-phytase clones (D'Alessio *et al.*, 1992, *Focus*® 14: 76) yielding *E. coli* DH5 α (pMWR46).

5

Example 4: DNA Sequencing of *Thermomyces lanuginosus* CBS 586.94 Phytase Gene

Restriction mapping of the pZL1-phytase clone *E. coli* DH5 α (pMWR46) described in Example 3 was performed using standard methods (Maniatis *et al.*, 1982, *supra*). DNA sequencing of the phytase clones was performed with an Applied Biosystems Model 373A Automated DNA Sequencer (Applied Biosystems, Inc., Foster City, CA) using the primer walking technique with dye-terminator chemistry (Giesecke *et al.*, 1992, *Journal of Virol. Methods* 38: 47-60). In addition to the lac-forward and lac-reverse primers, specific oligonucleotide sequencing primers were synthesized on an Applied Biosystems Model 394 DNA/RNA Synthesizer (Applied Biosystems, Inc., Foster City, CA) according to the manufacturer's instructions.

10

15

Example 5: Properties of the *Thermomyces lanuginosus* CBS 586.94 Phytase Gene

20

DNA sequencing of a portion of a cloned *Thermomyces lanuginosus* CBS 586.94 phytase gene (*E. coli* DH5 α - pMWR46) demonstrated an open reading frame (SEQ ID NO:1) (Figure 2) with homology to the *Aspergillus niger* (*ficum*) NRRL 3135 phytase gene (Figure 3).

25

30

The positions of introns and exons within the *Thermomyces lanuginosus* CBS 586.94 phytase gene were assigned based on alignments of the deduced amino acid sequence to the deduced amino acid sequence of the corresponding *Aspergillus niger* (*ficum*) NRRL 3135 phytase gene product. On the basis of this comparison, the *Thermomyces lanuginosus* CBS 586.94 phytase gene is comprised of 2 exons (47 and 1377 bp) which are interrupted by 1 small intron (56 bp). The size and composition of the intron is consistent with those of other fungal genes (Gurr *et al.*, 1987, In Kinghorn, J.R. (ed.), *Gene Structure in Eukaryotic Microbes*, pp. 93-139, IRL Press, Oxford) in that all contain consensus splice donor and acceptor sequences as well as the consensus lariat sequence (PuCTPuAC) near the 3' end of each intervening sequence.

The deduced amino acid sequence of the *Thermomyces lanuginosus* CBS 586.94 gene product is shown in Figure 2 (SEQ ID NO:2). Based on the rules of von Heijne (1984, *Journal of Molecular Biology* 173: 243-251), the first 22 amino acids of the *Thermomyces lanuginosus* gene product likely comprise a secretory signal peptide which directs the nascent polypeptide into the endoplasmic reticulum. The mature phytase is an acidic protein (predicted isoelectric point = 5.4) composed of 452 amino acids (MW = 51 kDa). The deduced amino acid sequence contains the active site motif RHGXRXXP (SEQ ID NO:2) which is shared by other known fungal phytases (Ullah and Dischinger, 1993, *Biochem. Biophys. Res. Commun.* 192: 754-759).

The deduced amino acid sequence of the mature *Thermomyces lanuginosus* CBS 586.94 phytase shares approximately 47.5% identity with the phytase from *Aspergillus niger* (ficiuum) NRRL 3135 as shown in Figure 3 (SEQ ID NO:3).

Example 6: Cross-Hybridization Studies Using Genomic DNA from Other Fungi

The cloned *Thermomyces lanuginosus* CBS 586.94 phytase gene was used as a probe in Southern hybridization experiments with genomic DNA samples from a variety of fungal genera. Southern blots were probed under conditions of low stringency (25% formamide, 5 x SSPE, 0.3% SDS at 42°C) and medium stringency (35% formamide, 5 x SSPE, 0.3% SDS at 42°C). Genomic DNA samples were isolated from the following species using the protocol outlined in Example 1: *Corynascus thermophilus* (ATCC 22066), *Fusarium graminearum* (ATCC 20334), *Humicola grisea* var. *thermoidea* (ATCC 16453), *Neurospora crassa* (FGSC 987), *Botrytis cinerea* (ATCC 11542), *Curvularia verruculosa* (CBS 147.63), *Rhizoctonia solani* (IMI 358730) *Trichoderma harzianum* (CBS 819.68), *Absidia sporophora-variabilis* (ATCC 36019), *Myceliophthora thermophila* (CBS 117.65), and *Penicillium diversum* (CBS 320.48). Each DNA sample (ca. 5 µg) was digested with *Bam*HI prior to electrophoresis on a 1% agarose gel. The DNA was blotted to Zeta-Probe nylon membrane (BioRad Laboratories, Hercules, CA) and probed with a nick translated DNA probe comprising the *phyI* gene. The blots were washed with 2 x SSPE ± 0.1% SDS at 42°C.

The phytase gene from *Thermomyces lanuginosus* CBS 586.94 cross-hybridized with probable phytase gene sequences in several other fungal species (Table 1). Under conditions of low stringency strong hybridization signals were apparent in DNAs from *Corynascus thermophilus*, *Fusarium graminearum*, *Humicola grisea* var. *thermoidea*, *Neurospora crassa*

and, of course, *Thermomyces lanuginosus*. Weaker signals were detected in *Botrytis cinerea*, *Curvularia verruculosa*, *Rhizoctonia solani*, and *Trichoderma harzianum*. No hybridization was detected in *Absidia sporophora-variabilis*, *Myceliophthora thermophila*, or *Penicillium diversum*. Using medium stringency, strong hybridization signals were visible with only 5 *Corynascus thermophilus*, *Fusarium graminearum*, and *Thermomyces lanuginosus*. Weak hybridization was observed with DNAs from *Humicola grisea* var. *thermoidea* and *Neurospora crassa*. These data indicate that the *Thermomyces lanuginosus* CBS 586.94 phytase gene can be used as a probe to clone phytase genes from other filamentous fungi.

10 **Table 1.** Hybridization of genomic DNA samples from various fungi probed with the cloned *Thermomyces lanuginosus* phytase gene. A +++ denotes a strong positive hybridization signal, + denotes a weak signal, and, - denotes no detectable hybridization.

	Genomic DNA source	Low Stringency	High Stringency
15	<i>Absidia sporophora-variabilis</i>	-	-
	<i>Botrytis cinerea</i>	+	-
	<i>Corynascus thermophilus</i>	+++	+++
	<i>Curvularia verruculosa</i>	+	-
	<i>Fusarium graminearum</i>	+++	+++
20	<i>Humicola grisea</i> var. <i>thermoidea</i>	+++	+
	<i>Myceliophthora thermophila</i>	-	-
	<i>Neurospora crassa</i>	+++	+
	<i>Penicillium diversum</i>	-	-
	<i>Rhizoctonia solani</i>	+	-
25	<i>Thermomyces lanuginosus</i>	+++	+++
	<i>Trichoderma harzianum</i>	+	-

Example 7: Construction of the Phytase Expression Vector pMWR48

Two synthetic oligonucleotide primers shown below were designed to PCR amplify 30 the *Thermomyces lanuginosus* CBS 586.94 phytase gene coding sequence from plasmid pMWR46 (*E. coli* DH5 α -pMWR46) for subcloning and expression in a *Fusarium* host.

Forward primer: 5'-ATTAAATGGCGGGGATAGTTGG-3' (SEQ ID NO:4)

Reverse primer: 5'-CTTAATTAATCAAAAGCAGCGATCCC-3' (SEQ ID NO:5)

The sense primer was designed to the first in-frame ATG and extends 14 bp downstream.

The antisense primer was designed to a region 14 bp upstream of the translational stop codon and extends through the stop codon.

In order to facilitate the subcloning of the gene fragment into an expression vector designated pDM181 (Figure 4), *Swal* and *PacI* restriction enzyme sites were introduced at the 5' and 3' end of the phytase gene, respectively. The vector pDM181 contained the *Fusarium oxysporum* trypsin-like protease promoter and terminator (WO 96/00787) as regulatory sequences. The plasmid also contained the *bar* gene as a selectable marker for fungal transformations (de Block *et al.*, 1987, *EMBO Journal* 6:2513-2518).

One hundred picomoles of each of the primers above were used in a PCR reaction containing 52 ng of pEJG13, 1X *Pwo* Buffer (Boehringer Mannheim, Indianapolis, IN), 1 mM each dATP, dTTP, dGTP, and dCTP, and 2.5 units of *PwoI* (Boehringer Mannheim, Indianapolis, IN). The amplification conditions were one cycle at 94°C for 2 minutes, 50°C for 30 seconds, and 72°C for 1 minute; 9 cycles each at 94°C for 15 seconds, 50°C for 30 seconds, and 72°C for 1 minute; 15 cycles each at 94°C for 15 seconds, 55°C for 30 seconds, and 72°C for 1 minute plus 20 seconds for each additional cycle; one cycle at 94°C for 15 seconds, 55°C for 30 seconds, and 72°C for 7 minutes; and a soak cycle at 4°C. The amplified 2866 bp DNA fragment was purified by gel electrophoresis and cut with restriction endonucleases *Swal* and *PacI* (using conditions specified by the manufacturer). The cut fragment was cloned into pDM181 (Figure 4) that had been previously cut with *Swal* and *PacI* resulting in the expression plasmid pMWR48 (Figure 5) in which transcription of the phytase gene was under the control of the *Fusarium oxysporum* trypsin-like protease promoter. The plasmid pMWR48 was transformed into *E. coli* DH5 α cells. The *E. coli* transformant containing the pMWR48 plasmid was isolated and plasmid DNA was prepared according to procedures described by Sambrook *et al.*, 1989, *supra*.

Example 8: Transformation of *Fusarium* CC1-3 and Analysis of Transformants

Fusarium strain CC1-3, a highly branched morphological mutant of *Fusarium* strain A3/5 (ATCC 20334) (Wiebe *et al.*, 1992, *Mycological Research* 96: 555-562; Wiebe *et al.*, 1991, *Mycological Research* 95: 1284-1288; Wiebe *et al.*, 1991, *Mycological Research* 96:

555-562), was grown in a liquid medium containing Vogel's salts, (Vogel, 1964, *Am. Nature* 98: 435-446), 25 mM NaNO₃, and 1.5% glucose for 4 days at 28°C and 150 rpm. Conidia were purified by filtration through 4 layers of cheesecloth and finally through one layer of Miracloth. Conidial suspensions were concentrated by centrifugation. Fifty ml of YPG medium comprised of 1% yeast extract, 2% bactopeptone, and 2% glucose were inoculated with approximately 10⁸ conidia, and incubated for 14 hours at 24°C and 150 rpm. Resulting hyphae were trapped on a sterile 0.4 µm filter and washed successively with sterile distilled water and 1.0 M MgSO₄. The hyphae were resuspended in 10 ml of NOVOZYM 234™ solution (2-10 mg/ml in 1.0 M MgSO₄) and digested for 15-30 minutes at 34°C with agitation at 80 rpm. Undigested hyphal material was removed from the resulting protoplast suspension by successive filtration through 4 layers of cheesecloth and through Miracloth. Twenty ml of 1 M sorbitol were combined with the protoplast solution. After mixing, the protoplasts were pelleted by centrifugation and washed successively by resuspension and centrifugation in 20 ml of 1 M sorbitol and in 20 ml of STC (0.8 M sorbitol, 0.05 M Tris pH 8.0, 0.05 M CaCl₂). The washed protoplasts were resuspended in 4 parts STC and 1 part SPTC (0.8 M sorbitol, 40% PEG 4000, 0.05 M Tris pH 8.0, 0.05 M CaCl₂) at a concentration of 5 x 10⁷/ml. One hundred µl of protoplast suspension were added to 5 µg of pMWR48 in polypropylene tubes (17 x 100 mm), mixed and incubated on ice for 30 minutes. One ml of SPTC was mixed gently into the protoplast suspension and incubation was continued at room temperature for 20 minutes. 12.5 ml of molten solution (cooled to 40°C) consisting of 1X Vogel's salts, 25 mM NaNO₃, 0.8 M sucrose and 1% low melting agarose (Sigma Chemical Company, St. Louis, MO) were mixed with the protoplasts and then plated onto an empty 100 mm Petri plate. Incubation was continued at room temperature for 10 to 14 days. After incubation at room temperature for 24 hours, 12.5 ml of the identical medium plus 10 mg of basta (Hoechst Schering, Rodovre, Denmark) per ml were overlayed onto the Petri plate. Basta was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1), and once with chloroform:isoamyl alcohol (24:1) before use. After two weeks, 17 transformants were apparent. A mycelial fragment from the edge of each transformant was transferred to individual wells of a 24 well plate containing Vogel's/BASTA medium. The medium contained 25 g of sucrose, 25 g of Noble agar, 20 ml of 50X Vogel's salts (Vogel, 1964, *supra*), 25 mM NaNO₃, and 10 g of basta

per liter. The plate was sealed in a plastic bag to maintain moisture and incubated approximately one week at room temperature.

Example 9: Expression of *Thermomyces lanuginosus* CBS 586.94 Phytase Gene

5 A mycelial fragment from each of the 17 *Fusarium* CC1-3 transformants described in Example 8 was inoculated into 20 ml of M400Da medium containing 50 g of maltodextrin, 2.0 g of MgSO₄·7H₂O, 2.0 g of KH₂PO₄, 4.0 g of citric acid, 8.0 g of yeast extract, 2.0 g of urea, and 0.5 ml of trace metals solution per liter and incubated for 7 days at 30°C and 150 rpm. The medium was adjusted to pH 6.0 with 5 N NaOH. The trace metals solution
10 contained 14.3 g of ZnSO₄·7H₂O, 2.5 g of CuSO₄·5H₂O, 0.5 g of NiCl₂·6H₂O, 13.8 g of FeSO₄·7H₂O, 8.5 g of MnSO₄·H₂O, and 3.0 g of citric acid per liter. The untransformed host was also run as a control. One ml of culture supernatant was harvested at 4, 5, and 7 days and stored at 4°C. Phytase activity was determined as described below.

15 A sample was diluted in 0.2 M citrate pH 5.5 buffer and 1 ml of the diluted sample was added to each of two test tubes. Two milliliters of 15% trichloroacetic acid (TCA) were added to one tube, while the other tube was pre-incubated at 40°C for 5 min. One milliliter of 1% phytate substrate solution in 0.2 M citrate pH 5.5 buffer was subsequently added to both tubes. Samples lacking TCA were incubated at 40°C for 30 minutes. At the end of the incubation period, 2 ml of TCA were added and the samples were allowed to cool. A
20 volume of 100 microliters of each sample was then diluted to 1 ml in water and pre-incubated at 50°C for 5 minutes. One milliliter of a reagent containing 1:2:1:1 of 6 N sulfuric acid:water:2.5% ammonium molybdate:10% ascorbic acid was added to each sample and the samples were incubated another 15 minutes for color development. The resulting color was measured at 690 nm on a Molecular Devices Thermomax Microplate Reader (Molecular
25 Devices, Sunnyvale, CA).

Spores from primary transformants producing the highest phytase activity were generated by inoculating 20 ml of a medium containing per liter 12.1 g of NaNO₃, 50 g of succinic acid, and 20 ml of 50X Vogel's salts (adjusted to pH 6.0) with mycelia and incubating at 30°C with shaking for 2-3 days. Single spores were isolated by spreading 150 ml of spore culture onto micromanipulator plates containing 1X Vogel's salts, 25 mM NaNO₃, 2.5% sucrose, 2% Noble agar and 5 mg/ml BASTA and using a micromanipulator to transfer single spores to a clear region of the plate. After 3 days growth at room
30

temperature, the germinated spores were transferred to individual Vogel plates containing 5 mg/ml BASTA.

Culture supernatants from fourteen of the seventeen primary transformants of pMWR48 were positive when assayed for phytase activity. Two primary transformants designated transformant #3 and #5 were selected for single spore isolation based on phytase activity. Nine single spore isolates were obtained. Shake flasks containing 25 ml of M400Da medium plus 5 mg/ml BASTA were inoculated in duplicate with mycelial chunks from each single spore isolate and incubated at 30°C. One ml aliquots of the culture media were harvested at four, five, and seven days post-inoculation and assayed for phytase activity. The results of the phytase assays demonstrated that both primary transformants produced activity.

Example 10: Production of recombinant *Thermomyces lanuginosus* CBS 586.94 Phytase

The primary *Fusarium* transformant #5 described in Example 9 was cultivated in two 2 liter fermentors for 7 days at 30°C in a medium at pH 6.25 comprised of 20 g of soy, 20 g of glucose, 10 g of yeast extract, 2 g of MgSO₄·7H₂O, 2 g of KH₂PO₄, 2 g of citric acid, 3 g of K₂SO₄, 2 g of CaCl₂·2H₂O, and 0.5 ml of trace metals solution per liter and fed with a medium comprised of 300 g of glucose, 20 g of (NH₄)₂HPO₄, and 1 g of citric acid per liter. A 2.5% inoculum was used for the fermentations. The inoculum was a 48-72 hours shake flask culture comprised of 62.5 g of Nutriose, 2 g of MgSO₄·7H₂O, 10 g of KH₂PO₄, 2 g of K₂SO₄, 2 g of citric acid, 10 g of yeast extract, 2 g of urea, 0.5 g of CaCl₂, and 0.5 ml of trace metals solution pH 6.0 per liter.

The whole culture broth was filtered using a double layer of Miracloth affixed to the top of a 4 liter beaker with rubber bands. The filtrate was recovered and then frozen at -20°C.

Example 11: Purification of Recombinant *Thermomyces lanuginosus* CBS 586.94 Phytase

The frozen cell-free broth (1700 ml) described in Example 9 was thawed, clarified by centrifugation at 10,000 x g, and concentrated to a volume of 350 ml with an Amicon hollow fiber filtration unit equipped with an Amicon S1Y10 filter (Amicon, Beverly, MA). The sample was adjusted to pH 7, diluted to a conductivity of 2 mS, and loaded at room temperature onto a 75 ml bed volume Pharmacia Q-Sepharose Big Beads column (Pharmacia

Biotech, Uppsala, Sweden) pre-equilibrated with 20 mM Tris-Cl pH 7 buffer. The column was eluted at a flow rate of 5 ml per minute using the equilibration buffer until the effluent A₂₈₀ had decreased to near baseline. The column was then eluted with a 600 ml gradient of 0-0.6 M NaCl in the same buffer at a flow rate of 5 ml per minute. Phytase enzyme activity was determined by measuring the hydrolysis rate of 10 mM *p*-nitrophenyl phosphate in 0.2 M sodium citrate pH 5.5 buffer at 405 nm and 30°C in a 200 µl reaction volume using a Molecular Devices Thermomax Microplate Reader. The bound enzyme activity eluted in fractions corresponding to ca. 0.2 M NaCl.

The active fractions were pooled and concentrated by ultrafiltration with an Amicon PM-10 membrane (Amicon, Beverly, MA) to a volume of 25 ml, diluted to a conductivity of 0.9 mS, and loaded at 4 ml per minute onto a Pharmacia MonoQ HR 10/16 column (Pharmacia Biotech, Uppsala, Sweden) pre-equilibrated in 20 mM MOPS pH 7 buffer. The column was eluted with 80 ml of the equilibration buffer and then with a 400 ml gradient of 0-0.5 M NaCl in the same buffer. Enzyme activity was detected in fractions using the *p*-nitrophenyl phosphate assay described above. The active fractions were also analyzed with a Novex 10-27% gradient SDS-polyacrylamide gel according to the manufacturer's instructions (Novex, San Diego, CA) and the fractions were pooled if judged by electrophoresis to be substantially purified.

The pooled fractions were concentrated with an Amicon PM-10 membrane by ultrafiltration and exchanged into 20 mM MES pH 5.5 buffer. The sample conductivity was 1.1 mS. One third of this sample was loaded at 1 ml per minute onto a Pharmacia Mono S HR 5/5 column (Pharmacia Biotech, Uppsala, Sweden) pre-equilibrated with 20 mM MES pH 5.5 buffer. The column was eluted with 5 ml of the equilibration buffer and then with a 25 ml linear gradient of 0-0.6 M sodium chloride in the same buffer. The active fractions were pooled after electrophoretic analysis to eliminate those which contained trace contaminants.

The purified recombinant *Thermomyces lanuginosus* CBS 586.94 phytase appeared homogeneous by SDS-polyacrylamide gel electrophoresis analysis with a single component that had a molecular weight of about 60,000 daltons.

Example 12: Physicochemical characterization of the purified recombinant *Thermomyces lanuginosus* CBS 586.94 phytase

All characterizations were performed with the purified recombinant *Thermomyces lanuginosus* CBS 586.94 phytase described in Example 11. Comparisons were also made where noted with a standard of an *Aspergillus niger (ficiuum)* phytase obtained from Novo Nordisk A/S, Bagsværd, Denmark.

Isoelectric Point. The isoelectric point (pI) of the *Thermomyces* phytase was determined to the *Aspergillus niger* phytase. Isoelectric focusing (IEF) was performed with a Novex pH 3-7 IEF gel (Novex, San Diego, CA) according to the manufacturer's instructions. IEF standards from both Pharmacia (Pharmacia, Uppsala, Sweden) and BioRad (BioRad Laboratories, Hercules, CA) were used to calibrate the gel.

IEF demonstrated that the *Thermomyces* phytase contained multiple components with pIs in the range of 4.7 to 5.2, whereas the *Aspergillus niger* phytase was observed to possess a pI of 4.8-5.0.

Amino terminal sequence analysis. The purified recombinant *Thermomyces* phytase was subjected to amino terminal protein sequence analysis on an Applied Biosystems Model 476A Sequencer (Applied Biosystems, Inc., Foster City, CA) according to the manufacturer's instructions with liquid phase TFA delivery and on-line HPLC for PTH-amino acid separation. Samples were spotted onto a Biobrene™ coated TFA glass fiber filter (Applied Biosystems, Inc., Foster City, CA) and sequenced directly.

Amino terminal sequence analysis of the purified phytase revealed three components. The major component (ca. 60%) is H₂N-His-Pro-Asn-Val-Asp-Ile-Ala-Arg-His-Trp-Gly-Gln-Tyr-Ser-Pro-Phe-Phe-Ser-Leu-Ala (SEQ ID NO:2) which corresponded to a kex2 cleavage site at position 34 in the primary translation product. Two minor sequences, (ca. 30%) H₂N-Gly-Glu-Asp-Glu-Pro-Phe-Val-Arg-Val-Leu-Val-Asn-Asp-Arg-Val-Val-Pro-Leu-His-Gly (SEQ ID NO:2) and (ca. 10%) H₂N-Ser-Glu-Glu-Glu-Glu-Gly-Glu-Asp-Glu-Pro-Phe-Val-Arg-Val-Leu-Val-Asn-Asp-Arg (SEQ ID NO:2) corresponded to internal cleavage sites near the COOH terminal of the protein at positions 428 and 435 in the primary translation product.

Enzyme kinetics studies. Enzyme kinetics studies were performed on the *Thermomyces* phytase and the *Aspergillus niger* phytase. The studies were accomplished by assay of inorganic phosphate liberated from corn phytic acid (Sigma Chemical Co., St.

Louis, MO). Standard enzyme reactions were carried out for 30 minutes at 37°C and pH 5.5 in 0.5% w/w phytic acid. The reaction was quenched by the addition of an equal volume of 15% w/w trichloroacetic acid. After cooling, 100 μ l of the resulting mixture was diluted into 1.0 ml of glass distilled water. The sample was incubated at 50°C for 5 minutes. Color reagent (1.0 ml) was added and the 50°C incubation was continued for 15 minutes. The absorbance of a 200 μ l aliquot was measured at 690 nm with a Molecular Devices Thermomax Microplate Reader. The color reagent is composed of 6 N sulfuric acid: water: 2.5% (w/v) ammonium heptamolybdate, tetrahydrate: 10% ascorbate (aqueous) in a ratio of 1:2:1:1 and prepared fresh on a daily basis. Quantitation was based on a standard curve generated with a 10 mM sodium monobasic phosphate standard. One unit (U) is defined as a μ mole of inorganic phosphate released per minute at 37°C and pH 5.5.

Steady-state kinetics measurements were made by substrate titration. Phytate concentrations were 2.16, 1.08, 0.541, 0.216, 0.108, and 0.0758 mM for the purposes of K_m determination. Phytate concentrations of 1.08, 0.541, 0.216 and 0.108 mM in the presence or absence of 1 mM sodium monobasic phosphate were used for the purpose of evaluating product inhibition.

Steady-state kinetic measurements demonstrated that the *Thermomyces* phytase has an apparent K_m of approximately 110 μ M with respect to phytate while *Aspergillus niger* phytase has an apparent K_m of 200 μ M. There was a weak indication of excess substrate inhibition at the 2.16 mM substrate concentration, perhaps congruent with the literature report of inhibition above 2 mM for *Aspergillus* phytase (Ullah, 1988, *Preparative Biochemistry* 18: 443-458). Steady-state kinetic measurements with 1 mM phosphate present failed to reveal any type of inhibition with this product. It was estimated that the K_i for phosphate must exceed 3 mM to be undetectable in our experiments. In contrast, Ullah (Ullah, 1988, *supra*) has reported that phosphate is a competitive inhibitor of *Aspergillus* phytase with a K_i of 1.9 mM.

Thermostability measurement. The thermostability of the *Thermomyces* phytase was compared to the *Aspergillus niger* phytase. Samples of the *Thermomyces* phytase and the *Aspergillus niger* phytase were dissolved at 100 U per ml in 0.2 M sodium citrate pH 5.5 buffer. Aliquots (100 μ l) of each enzyme solution were incubated for 20 minutes in a water bath at the following temperatures: 37, 45, 50, 55, 60, 65, 70 and 75°C. After the temperature treatments, the samples were stored at 0°C until activity assays were performed.

Each sample was diluted 1:80 into 0.2 M sodium citrate pH 5.5 buffer containing 0.01% w/w Tween 20 and the standard activity assay described above was performed.

Comparison of the enzyme thermostability profiles (Figure 6) suggested that stability differences between the two enzymes are small. However, neither enzyme is fully inactivated by a high temperature incubation and the residual activity profile is consistent with partially reversible thermal denaturation (Ullah and Mullaney, 1996, *Biochem. Biophys. Res. Comm.* 227: 311-317).

pH-activity measurement. The pH-activity profile of the *Thermomyces* phytase was compared to the *Aspergillus niger* phytase. To attain a buffering range between pH 2-7, a three component 0.125 M glycine-acetate-citrate buffer was employed. The buffer components were combined at final concentrations of 42 mM per component and phytic acid was added as a solid to 1% w/w. This mixture was adjusted to pH 7 with concentrated HCl and a 10 ml aliquot was taken out. This was done for every 0.5 pH increment thereafter through pH 2.

Enzyme stock solutions of 20 U per ml were prepared in 20 mM MES buffer pH 5.5. Substrate (850 microliters) in buffer at a given pH was combined with 100 microliters of water and 50 microliters of enzyme stock solution and incubated for 30 minutes at 37°C. Subsequently the enzyme reaction was quenched with 1 ml of 15% TCA and quantitated by the standard method.

The pH-activity profile comparison of *Thermomyces* and *Aspergillus* phytase indicated substantial similarity between the two enzymes in pH profile (Figure 7). However, the *Thermomyces* phytase is also active at neutral pH while the *Aspergillus* enzyme is not. Earlier reports in the literature (see, for example, Sandberg, *et al.*, 1996, *J. Nutr.* 126: 476-480) suggest that *Aspergillus* phytase possesses two pH optima. It is more likely that these reports are based on impure material which contains traces of the *Aspergillus* acid phosphatase which is well-known in the literature (Zyla, 1993, *World J. Microbiol. Biotechnol.* 9: 117-119).

Temperature-activity measurement. The temperature-activity profile of the *Thermomyces* phytase was compared to the *Aspergillus niger* phytase. Enzyme stock solutions of 12.5 U per ml were prepared in 0.2 M sodium citrate pH 5.5 buffer. 250 microliters of 1% phytic acid substrate was added to a 1.7 ml Eppendorf tube followed by 240 microliters of 0.2 M sodium citrate pH 5.5 buffer. This solution was vortexed and

placed in a water bath at the designated temperature. After a 20 minute equilibration in the water bath, the Eppendorf tube was vortexed and 10 microliters of phytase solution was added. The sample was vortexed and incubated in the water bath for an additional 30 minutes. After 30 minutes in the water bath, the reaction was quenched with 1 ml of 15% TCA and quantitated by the assay method described above.

Measurement of enzyme activity as a function of temperature revealed a significant difference between the two enzymes (Figure 8). *Thermomyces* phytase displays maximum enzyme activity near 65°C and has partial activity even at 75°C while the *Aspergillus niger* phytase is essentially inactive at 65°C.

Phytate Hydrolysis. A comparison of the ability of *Thermomyces* and *Aspergillus niger* phytases to hydrolyze phytic acid was made. Each phytase (0.5 U enzyme activity per ml) was incubated with either 0.5% or 0.1% phytic acid at pH 5.5 and 37°C for 10 hours. The results revealed that *Aspergillus niger* and *Thermomyces* phytases released identical amounts (70%) of the total theoretically available phosphorus after 10 hours with either 0.5% or 0.1% phytic acid concentrations.

Deposit of Biological Materials

The following strain has been deposited according to the Budapest Treaty in the Agricultural Research Service Patent Culture Collection (NRRL), Northern Regional Research Laboratory, 1815 University Street, Peoria, Illinois 61604, USA.

Strain	Accession Number	Deposit Date
<i>E. coli</i> DH5 α (pMWR46)	NRRL B-21527	February 23, 1996

The strain has been deposited under conditions that assure that access to the culture will be available during the pendency of this patent application to one determined by the Commissioner of Patents and Trademarks to be entitled thereto under 37 C.F.R. §1.14 and 35 U.S.C. §122. The deposit represents a substantially pure culture of each deposited strain. The deposit is available as required by foreign patent laws in countries wherein counterparts of the subject application, or its progeny are filed. However, it should be understood that the availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by governmental action.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations

of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the
5 appended claims.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT:

- (A) NAME: Novo Nordisk Biotech, Inc.
- (B) STREET: 1445 Drew Avenue
- (C) CITY: Davis
- (D) STATE: California
- (E) COUNTRY: US
- (F) ZIP: 95616-4880
- (G) TELEPHONE: (916) 757-8100
- (H) TELEFAX: (916) 758-0317

(ii) TITLE OF THE INVENTION: Polypeptides Having Phytase Activity
And Nucleic Acids Encoding Same

(iii) NUMBER OF SEQUENCES: 5

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Novo Nordisk of North America, Inc.
- (B) STREET: 405 Lexington Avenue, Suite 6400
- (C) CITY: New York
- (D) STATE: New York
- (E) COUNTRY: U.S.A.
- (F) ZIP: 10174-6401

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette
- (B) COMPUTER: IBM Compatible
- (C) OPERATING SYSTEM: DOS
- (D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: to be assigned
- (B) FILING DATE: to be assigned
- (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Lambiris, Elias J.
- (B) REGISTRATION NUMBER: 33,728
- (C) REFERENCE/DOCKET NUMBER: 4758.200-WO

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 212 867 0123
- (B) TELEFAX: 212 867 0298

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2200 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATTACGGAG TAGTTGCCAT TCGATGTTCA TTGATCAACA GTCAACCGCA AGTTTCGTAG	60
TATTTTCCAA ACTCCTCCAC TGGCCGTGCG TTGCGACACG ACCTGCATGA GAATCGATCG	120
ATCGATCGAT GATCGCTCAG GATGATCTGA TCATCTCGGG TTGGAAGAGT CCACTTTATG	180
ACCAGGGGAT TGATTTTCA ATGCGTTGGT TGTTGTTCAT CCGATTCAAG AACAAAGTGGA	240

CATTATTATT ATGATTGCAC GTGTCTTAAG CTGCAAGTAC TATTGAATAG TGCTTCAATG	300
CTACCATGAT CGGACACCAA CACTCATGGA AGCCCGCCCC TAGCCGGCAG ATCTGGCACA	360
CGCATCGTGC TGATATAAAA AGACTGCCAA ATGCCGAAGA CGAAATGCAG CAACGTTCAG	420
CCCGCAGAGT GATTGCCGTC ATGGCGGGGA TAGGTTGGG GTCCTTCTG GTCCTGCTGC	480
TGCAATTGTA CGCATTCTTC TAGACCCTAA TTATAGAGGT CTGTTGCTGA TATTCTGACT	540
AGTTCGGCAT TATTGACGGC CTCGCCGGCC ATTCCCTCCTT TCTGGAGGAA GAAGCATCCC	600
AACGTGGACA TTGCCCCCCA CTGGGGCCAG TACTCGCCCT TCTTCTCGCT GGCGAGGTC	660
TCTGAAATCT CGCCTGCGGT GCCCAAGGGC TGTCGTGTCG AGTTGTGCA GGTGCTGTCC	720
CGGCACGGAG CTCGGTATCC TACTGTCAC AAGAGTGAAG TCTACGCCGA GTTGCTTCAA	780
AGGATCCAGG ACACTGCGAC CGAGTTCAAG GGCGATTTG CCTTTCTCCG AGACTATGCC	840
TATCATCTCG GTGCCGATAA TTTGACGCGC TTTGGCGAGG AGCAGATGAT GGAATCGGGC	900
CGCCAGTTCT ACCACCGGTA TCGTGAGCAG GCCCGAGAGA TTGTGCCATT TGTGCGTGC	960
GCAGGGCTCCG CGCGAGTCAT TGCCTCGGCA GAGTTCTCA ACCGCGGATT CCAGGATGCC	1020
AAAGACCGGG ATCCCAGGTC GAACAAGGAC CAGGCAGAGC CTGTGATCAA CGTGATCATT	1080
TCCGAAGAAA CTGGCAGTAA CAATACTCTG GATGGGCTGA CGTGCCCCGC GGCGAGGAG	1140
GCACCGGACC CAACCCAGCC CGCAGAGTTC CTGCAAGTTT TC GGCCCCCG TGTTTGAAA	1200
AAGATCACTA AACACATGCC GGGTGTGAAC CTCACCTTGG AGGATGTCCC GTTGTTCATG	1260
GATCTTGTC CGTTGACAC GGTGGGCTCC GACCCAGTTC TTTTCCCACG GCAGCTCTCT	1320
CCGTTTGTC ACTTGTTCAC GGCGACGAT TGGATGGCCT ACGATTACTA CTACACCCCTC	1380
GACAAATACT ACAGCCACGG CGGCGGCAGC GCATTGGCC CGTCCCGCG CGTCGGGTT	1440
GTCAACGAGC TGATTGCGCG TATGACGGGA AATCTTCCCG TCAAGGACCA CACAACAGTC	1500
AACCACACTC TCGATGACAA CCCGGAAACT TTCCCGTTGG ACGCTGTCCT CTACGCAGAC	1560
TTTTCGCACG ACAACACCAT GACGGGCATC TTTCCGCAA TGGGCCTGTA CAACGGCACA	1620
AAGCCGCTGT CGACGTCCAA GATTCAACCT CCGACGGGTG CAGCAGCGGA TGGATATGCG	1680
GCATCGTGGAA CGGTGCCGTT CGCAGCGAGG GCGTATGTGG AGTTGCTGCG ATGTGAGACG	1740
GAAACGAGCT CTGAGGAGGA GGAGGAGGGGG GAGGACGAGC CGTTGCGTG GGTTCTGGTG	1800
AATGATCGGG TTGTGCCGCT GCATGGTTGT CGGGTTGATC GATGGGGGAG GTGTCGGAGG	1860
GATGAGTGGAA TTAAGGGACT CACGTTGCT CGACAGGGTG GGCATTGGGA TCGCTGCTTT	1920
TGATTAGATG CTCATAGACA TAACCCCATG ATTCCGAATT GATGTTTTA GATACAATCA	1980
CTGCGGAAAG GGAAATGATC CAAAAAGCGC CAGTCTAGTA TAACTTGCG AATCCGTTGA	2040
CTTGTTCAGT CCTTGGTGTC GCCATCAACC AGGCCTGCCA CAAGGTCCAA TGTTCCCGCT	2100
CTACATGGAG TCCGTCGTCG CCGAGATCAT CCACGCCAG CGCACGGAGC TGTTCCGTTG	2160
AGGGTATCTG CCGTGGTTGA CCCCCGTGCT CACAGTCACA	2200

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 475 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Gly Ile Gly Leu Gly Ser Phe Leu Val Leu Leu Leu Gln Phe
 1 5 10 15

Ser Ala Leu Leu Thr Ala Ser Pro Ala Ile Pro Pro Phe Trp Arg Lys
 20 25 30

Lys His Pro Asn Val Asp Ile Ala Arg His Trp Gly Gln Tyr Ser Pro
 35 40 45

Phe Phe Ser Leu Ala Glu Val Ser Glu Ile Ser Pro Ala Val Pro Lys
 50 55 60

Gly Cys Arg Val Glu Phe Val Gln Val Leu Ser Arg His Gly Ala Arg
 65 70 75 80

Tyr Pro Thr Ala His Lys Ser Glu Val Tyr Ala Glu Leu Leu Gln Arg
 85 90 95

Ile Gln Asp Thr Ala Thr Glu Phe Lys Gly Asp Phe Ala Phe Leu Arg
 100 105 110

Asp Tyr Ala Tyr His Leu Gly Ala Asp Asn Leu Thr Arg Phe Gly Glu
 115 120 125

Glu Gln Met Met Glu Ser Gly Arg Gln Phe Tyr His Arg Tyr Arg Glu
 130 135 140

Gln Ala Arg Glu Ile Val Pro Phe Val Arg Ala Ala Gly Ser Ala Arg
 145 150 155 160

Val Ile Ala Ser Ala Glu Phe Phe Asn Arg Gly Phe Gln Asp Ala Lys
 165 170 175

Asp Arg Asp Pro Arg Ser Asn Lys Asp Gln Ala Glu Pro Val Ile Asn
 180 185 190

Val Ile Ile Ser Glu Glu Thr Gly Ser Asn Asn Thr Leu Asp Gly Leu
 195 200 205

Thr Cys Pro Ala Ala Glu Glu Ala Pro Asp Pro Thr Gln Pro Ala Glu
 210 215 220

Phe Leu Gln Val Phe Gly Pro Arg Val Leu Lys Lys Ile Thr Lys His
 225 230 235 240

Met Pro Gly Val Asn Leu Thr Leu Glu Asp Val Pro Leu Phe Met Asp
 245 250 255

Leu Cys Pro Phe Asp Thr Val Gly Ser Asp Pro Val Leu Phe Pro Arg
 260 265 270

Gln Leu Ser Pro Phe Cys His Leu Phe Thr Ala Asp Asp Trp Met Ala
 275 280 285

Tyr Asp Tyr Tyr Tyr Thr Leu Asp Lys Tyr Tyr Ser His Gly Gly Gly
 290 295 300
 Ser Ala Phe Gly Pro Ser Arg Gly Val Gly Phe Val Asn Glu Leu Ile
 305 310 315 320
 Ala Arg Met Thr Gly Asn Leu Pro Val Lys Asp His Thr Thr Val Asn
 325 330 335
 His Thr Leu Asp Asp Asn Pro Glu Thr Phe Pro Leu Asp Ala Val Leu
 340 345 350
 Tyr Ala Asp Phe Ser His Asp Asn Thr Met Thr Gly Ile Phe Ser Ala
 355 360 365
 Met Gly Leu Tyr Asn Gly Thr Lys Pro Leu Ser Thr Ser Lys Ile Gln
 370 375 380
 Pro Pro Thr Gly Ala Ala Ala Asp Gly Tyr Ala Ala Ser Trp Thr Val
 385 390 395 400
 Pro Phe Ala Ala Arg Ala Tyr Val Glu Leu Leu Arg Cys Glu Thr Glu
 405 410 415
 Thr Ser Ser Glu Glu Glu Glu Gly Glu Asp Glu Pro Phe Val Arg
 420 425 430
 Val Leu Val Asn Asp Arg Val Val Pro Leu His Gly Cys Arg Val Asp
 435 440 445
 Arg Trp Gly Arg Cys Arg Arg Asp Glu Trp Ile Lys Gly Leu Thr Phe
 450 455 460
 Ala Arg Gln Gly Gly His Trp Asp Arg Cys Phe
 465 470 475

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 467 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly Val Ser Ala Val Leu Leu Pro Leu Tyr Leu Leu Ser Gly Val
 1 5 10 15
 Thr Ser Gly Leu Ala Val Pro Ala Ser Arg Asn Gln Ser Ser Cys Asp
 20 25 30
 Thr Val Asp Gln Gly Tyr Gln Cys Phe Ser Glu Thr Ser His Leu Trp
 35 40 45
 Gly Gln Tyr Ala Pro Phe Phe Ser Leu Ala Asn Glu Ser Val Ile Ser
 50 55 60
 Pro Glu Val Pro Ala Gly Cys Arg Val Thr Phe Ala Gln Val Leu Ser
 65 70 75 80
 Arg His Gly Ala Arg Tyr Pro Thr Asp Ser Lys Gly Lys Lys Tyr Ser
 85 90 95
 Ala Leu Ile Glu Glu Ile Gln Gln Asn Ala Thr Thr Phe Asp Gly Lys
 100 105 110

Tyr Ala Phe Leu Lys Thr Tyr Asn Tyr Ser Leu Gly Ala Asp Asp Leu
 115 120 125

Thr Pro Phe Gly Glu Gln Glu Leu Val Asn Ser Gly Ile Lys Phe Tyr
 130 135 140

Gln Arg Tyr Glu Ser Leu Thr Arg Asn Ile Val Pro Phe Ile Arg Ser
 145 150 155 160

Ser Gly Ser Ser Arg Val Ile Ala Ser Gly Lys Lys Phe Ile Glu Gly
 165 170 175

Phe Gln Ser Thr Lys Leu Lys Asp Pro Arg Ala Gln Pro Gly Gln Ser
 180 185 190

Ser Pro Lys Ile Asp Val Val Ile Ser Glu Ala Ser Ser Ser Asn Asn
 195 200 205

Thr Leu Asp Pro Gly Thr Cys Thr Val Phe Glu Asp Ser Glu Leu Ala
 210 215 220

Asp Thr Val Glu Ala Asn Phe Thr Ala Thr Phe Val Pro Ser Ile Arg
 225 230 235 240

Gln Arg Leu Glu Asn Asp Leu Ser Gly Val Thr Leu Thr Asp Thr Glu
 245 250 255

Val Thr Tyr Leu Met Asp Met Cys Ser Phe Asp Thr Ile Ser Thr Ser
 260 265 270

Thr Val Asp Thr Lys Leu Ser Pro Phe Cys Asp Leu Phe Thr His Asp
 275 280 285

Glu Trp Ile Asn Tyr Asp Tyr Leu Gln Ser Leu Lys Lys Tyr Tyr Gly
 290 295 300

His Gly Ala Gly Asn Pro Leu Gly Pro Thr Gln Gly Val Gly Tyr Ala
 305 310 315 320

Asn Glu Leu Ile Ala Arg Leu Thr His Ser Pro Val His Asp Asp Thr
 325 330 335

Ser Ser Asn His Thr Leu Asp Ser Ser Pro Ala Thr Phe Pro Leu Asn
 340 345 350

Ser Thr Leu Tyr Ala Asp Phe Ser His Asp Asn Gly Ile Ile Ser Ile
 355 360 365

Leu Phe Ala Leu Gly Leu Tyr Asn Gly Thr Lys Pro Leu Ser Thr Thr
 370 375 380

Thr Val Glu Asn Ile Thr Gln Thr Asp Gly Phe Ser Ser Ala Trp Thr
 385 390 395 400

Val Pro Phe Ala Ser Arg Leu Tyr Val Glu Met Met Gln Cys Gln Ala
 405 410 415

Glu Gln Glu Pro Leu Val Arg Val Leu Val Asn Asp Arg Val Val Pro
 420 425 430

Leu His Gly Cys Pro Val Asp Ala Leu Gly Arg Cys Thr Arg Asp Ser
 435 440 445

Phe Val Arg Gly Leu Ser Phe Ala Arg Ser Gly Gly Asp Trp Ala Glu
 450 455 460

Cys Phe Ala
465

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATTTAAATGG CGGGGATAGG TTTGG

25

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTTAATTAAT CAAAAGCAGC GATCCC

26

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13 bis)

A. The indications made below relate to the microorganism referred to in the description on page 41, lines 17-30.

B. IDENTIFICATION OF

Further deposits are identified on an additional sheet

Name of depository institution

Agricultural Research Service Patent Culture Collection (NRRL)

Address of depository institution (*including postal code and country*)

Northern Regional Research Center
1815 University Street
Peoria, IL 61604, US

Date of deposit
February 23, 1996

Accession Number
NRRL B-21527

C. ADDITIONAL INDICATIONS (*leave blank if not applicable*) This information is continued on an additional sheet

In respect of those designations in which a European and/or Australia Patent is sought, during the pendency of the patent application, a sample of the deposited microorganism is only to be provided to an independent expert nominated by the person requesting the sample (Rule 28(4) EPC/Regulation 3.25 of Australia Statutory Rule 1991 No. 71).

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (*if the indications are not for all designated States*)

E. SEPARATE FURNISHING OF INDICATIONS (*leave blank if not applicable*)

The indication listed below will be submitted to the International Bureau Later (*specify the general nature of the indications e.g. "Accession Number of Deposit"*)

For receiving Office use only

For International Bureau use only

This sheet was received with the international application

This sheet was received with the International Bureau on:

Authorized officer

Authorized officer

Claims

What is claimed is:

- 5 1. An isolated polypeptide having phytase activity, selected from the group consisting of:
 - (a) a polypeptide with an amino acid sequence which has at least 60% identity with the amino acid sequence set forth in SEQ ID NO:2;
 - (b) a polypeptide which is encoded by a nucleic acid sequence which is capable of hybridizing under medium stringency conditions with (i) the nucleic acid sequence set forth in SEQ ID NO:1, or (ii) its complementary strand;
 - (c) an allelic form of (a) or (b); and
 - (d) a fragment of (a), (b), or (c).
- 15 2. The polypeptide of claim 1, which has an amino acid sequence with at least 60% identity with the amino acid sequence set forth in SEQ ID NO:2 or a fragment thereof.
- 20 3. The polypeptide of claim 2, which has an amino acid sequence with at least 70% identity with the amino acid sequence set forth in SEQ ID NO:2 or a fragment thereof.
- 25 4. The polypeptide of claim 3, which has an amino acid sequence with at least 80% identity with the amino acid sequence set forth in SEQ ID NO:2 or a fragment thereof.
- 30 5. The polypeptide of claim 4, which has an amino acid sequence with at least 90% identity with the amino acid sequence set forth in SEQ ID NO:2 or a fragment thereof.
6. The polypeptide of claim 5, which has an amino acid sequence with at least 95% identity with the amino acid sequence set forth in SEQ ID NO:2 or a fragment thereof.
7. The polypeptide of claim 6, which has the amino acid sequence set forth in SEQ ID NO:2 or a fragment thereof.

8. The polypeptide of claim 7, which has the amino acid sequence set forth in SEQ ID NO:2.

9. The polypeptide of claim 2, which is obtained from a strain of *Thermomyces* or a synonym or teleomorph thereof.
5

10. The polypeptide of claim 9, which is obtained from a strain of *Thermomyces lanuginosus* or a synonym or teleomorph thereof.

10 11. The polypeptide of claim 10, which is obtained from *Thermomyces lanuginosus* CBS 586.94 or a mutant strain thereof.

12. The polypeptide of claim 1, which is encoded by a nucleic acid sequence which is capable of hybridizing under medium stringency conditions with (i) the nucleic acid sequence set forth in SEQ ID NO:1 or (ii) its complementary strand; or a fragment thereof.
15

13. The polypeptide of claim 12, which is encoded by a nucleic acid sequence which is capable of hybridizing under high stringency conditions with (i) the nucleic acid sequence set forth in SEQ ID NO:1 or (ii) its complementary strand; or a fragment thereof.

20 14. The polypeptide of claim 12, which is encoded by a nucleic acid sequence which is capable of hybridizing under medium stringency conditions with (i) the nucleic acid sequence set forth in SEQ ID NO:1 or (ii) its complementary strand.

25 15. The polypeptide of claim 12, which is obtained from a strain of *Thermomyces* or a synonym or teleomorph thereof.

16. The polypeptide of claim 15, which is obtained from a strain of *Thermomyces lanuginosus* or a synonym or teleomorph thereof.

30 17. The polypeptide of claim 1, which is encoded by the nucleic acid sequence contained in plasmid pMWR46 which is contained in *Escherichia coli* NRRL B-21527.

18. An isolated nucleic acid sequence comprising a nucleic acid sequence which encodes the polypeptide of claim 1.

19. The nucleic acid sequence according to claim 18, wherein the nucleic acid sequence encodes a polypeptide obtained from a *Thermomyces* strain.

20. The nucleic acid sequence according to claim 19, wherein the nucleic acid sequence encodes a polypeptide obtained from *Thermomyces lanuginosus* or a synonym or teleomorph thereof.

10

21. The nucleic acid sequence according to claim 20, wherein the nucleic acid sequence encodes a polypeptide obtained from *Thermomyces lanuginosus* CBS 586.94 or a mutant strain thereof.

15

22. The nucleic acid sequence according to claim 18, which encodes a polypeptide which has an amino acid sequence which has at least 60% identity with the amino acid sequence set forth in SEQ ID NO:2.

20

23. The nucleic acid sequence according to claim 18, which is capable of hybridizing under medium stringency conditions with (a) the nucleic acid sequence set forth in SEQ ID NO:1 or (b) its complementary strand; or an allelic form or fragment thereof.

25

24. The nucleic acid sequence according to claim 23, which is capable of hybridizing under high stringency conditions with (a) the nucleic acid sequence set forth in SEQ ID NO:1 or (b) its complementary strand.

25. The nucleic acid sequence according to claim 18, wherein the nucleic acid sequence is contained in plasmid pMWR46 which is contained in *Escherichia coli* NRRL B-21527.

30

26. The nucleic acid sequence according to claim 18, wherein the nucleic acid sequence is set forth in SEQ ID NO:1.

27. A nucleic acid construct comprising the nucleic acid sequence of claim 18 operably linked to one or more control sequences capable of directing the expression of the polypeptide in a suitable expression host.

5 28. A recombinant expression vector comprising the nucleic acid construct of claim 27, a promoter, and transcriptional and translational stop signals.

29. The vector according to claim 28, further comprising a selectable marker.

10 30. A recombinant host cell comprising the nucleic acid construct of claim 27.

31. The cell according to claim 30, wherein the nucleic acid construct is contained on a vector.

15 32. The cell according to claim 30, wherein the nucleic acid construct is integrated into the host cell genome.

33. The cell according to claim 30, wherein the host cell is a bacterial or fungal cell.

20 34. The cell according to claim 33, wherein the bacterial cell is a *Bacillus*, *Pseudomonas*, or *Streptomyces* cell.

35. The cell according to claim 33, wherein the fungal cell is a filamentous fungal or yeast cell.

25 36. The cell according to claim 35, wherein the filamentous fungal cell is a cell of a species of *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium*, or *Trichoderma*.

30 37. The cell according to claim 35, wherein the yeast cell is a cell of a species of *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* or *Yarrowia*.

38. A method for producing the polypeptide of claim 1 comprising (a) cultivating a *Thermomyces* strain to produce a supernatant comprising the polypeptide; and (b) recovering the polypeptide.

5 39. A method for producing the polypeptide of claim 1 comprising (a) cultivating a host cell comprising a nucleic acid construct comprising a nucleic acid sequence encoding the polypeptide under conditions conducive to expression of the polypeptide; and (b) recovering the polypeptide.

10 40. A feed or food composition comprising a polypeptide of claim 1 and one or more food additives or components.

41. A method for reducing phytate levels in animal manure comprising feeding an animal with an effective amount a feed or food composition of claim 40.

15 42. A method for liquefying a starch, comprising
 (a) treating the starch with a polypeptide of claim 1 prior to or simultaneously with liquefying; and
 (b) adding an α -amylase to the starch; and
20 (c) reacting the starch of step (b) for a time and at a temperature effective to liquefy the starch.

FIGURE 1

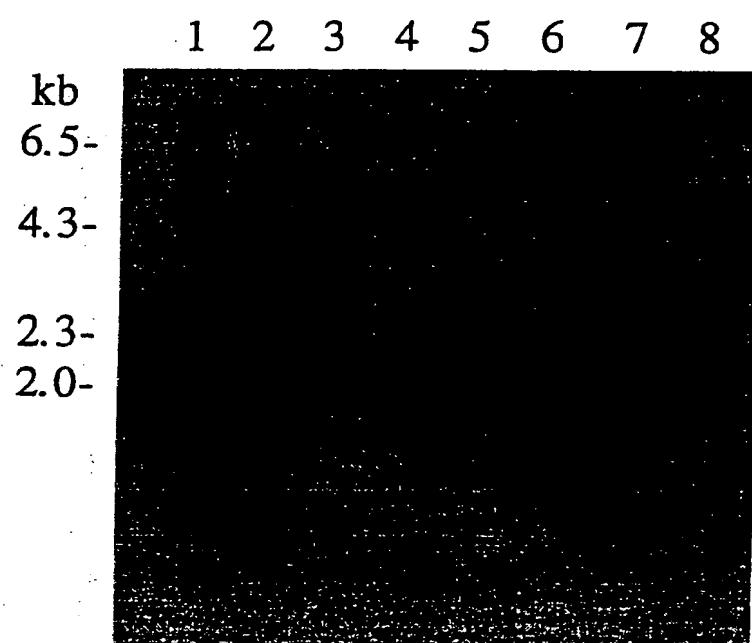


Figure 2

AATTACGGAGTAGTTGCCATTGGATGTCATTGATCAACAGTCACCGCAAGTTCTGTAGTATTTCCA
 90
 TTGGCAGACGGACCTGCATGAGAATCGATCGATCGATGATCGATCAGGATGATCTGATCATCTCGGGTGGAAAGAGTCCACTTATG
 180
 ACCAGGGGATTGATTTCAATGCGTGGTTGTTGTCATCCGATTCATGAGAACAGTOGGACATTATTATGATTGACCGTGCCAAAG
 270
 CTGCAAGTACTATTGAATAGTGCCTCAATGCTACCATGATCGACACCAACACTCATGGAAGCCGCCCCCTAGCCGGCAGATCTGGCACA
 360
 CGCATCGTGTGATATAAAAGACTGCCAAATGCCAGACGCCAAATGCCAGCAACGTTCAGCCCCAGAGTGAATTGCCGTATGGGGGGGA
 450
 M A G
 TAGGTTTGGGGTCTTCTGGTCTGCTGCAATTgtacgcattttctagaccctaattatagaggctgttgctgatattctgact
 540
 I G L G S F L V L L L Q F _____
 a9TCGGCATTATTGACGGCCTGGCGCCATTCCTCTTCTGGAGGAAGAACGATCCAACTGGGACATTGCCCGCCACTGGGGCCAG
 630
 — S A L L T A S P A I P P F W R K K H P N V D I A R H W G Q
 TACTCGCCCTTCTCTGCTGGCCGAGGTCTGAAATCTGCTGCGTGCCTAGGGCTGTGCTGCTGAGTTGTCAGGTGCTGTC
 720
 Y S P F F S L A E V S E I S P A V P K G C R V E F V Q V L S
 CGGCACGGAGCTGGTATCCTACTGTCACAGAGTGAAGTCTACGCCAGTTGCTCAAAGGATCCAGGACACTGGACCCAGTTCAAG
 810
 R H G A R Y P T A H K S E V Y A E L L Q R I Q D T A T E F K
 GGCATTTGCTTCTCCGAGACTATGCTATCATCTGCTGCGATAATTGACGCCCTTGGCGAGGAGCAGATGATGGAATGGGC
 900
 G D F A F L R D Y A Y H L G A D N L T R F G E E Q M M E S G
 CGCCAGTTCTACCACGGTATCTGAGCCAGGGAGAGATTTGCCATTGCGTGCAGGCTCCGCCAGTCATGCCGCA
 990
 R Q F Y H R Y R E Q A R E I V P F V R A A G S A R V I A S A
 GAGTTCTCAACCGCGGATCCAGGATGCCAAAGACGGGATCCAGGTGAAACAGGACAGGAGCTGTGATCAACGTGATCATT
 1080
 E F F N R G F Q D A K D R D P R S N K D Q A E P V I N V I
 TCCGAAGAAACTGGCAGTAACAAACTCTGGATGGGCTGACGTGCCCCGGGGAGGAGCCACGGGACCCAAAGGGCCAGAGTT
 1170
 S E E T G S N N T L D G L T C P A A E E A P D P T Q P A E F
 CTGCAAGTTTGGCCGGGTGCTTGAAGAAAGATCACTAAACACATGCCGGTGTGAACCTCACCTGGAGGATGTCCCCTGTCATG
 1260
 L Q V F G P R V L K K I T K H M P G V N L T L E D V P L F M
 GATCTTGTCCGTTGACACGGTGGCTCCGACCCAGTTCTTCCAGGGAGCTCTCCGGTTGTCACCTGTTCAAGGGGAGGAT
 1350
 D L C P F D T V G S D F V L F P R Q L S P F C H L F T A D D
 TGATGGCCCTACGATTACTACTACACCCCTGACAAAATACAGCCACGGGGCGAGGGCATTTGGCCCTCCGCCGGTGGGTT
 1440
 W M A Y D Y Y Y T L D K Y Y S H G G G S A F G P S R G V G F
 GTCAACGGAGCTGATTGCCGTATGCGGAAATCTTCCCGTCAAGGACCAACACAGTCACCCACACTCTGGATGACAAACGGGAACT
 1530
 V N E L I A R M T G N L P V K D H T T V N H T L D D N P E T
 TTCCCGTTGGACGCTGCTCTACGCAAGACTTTGCCACGACAAACACCATGACGGGATCTTCCCGCAATGGCCCTGTACAGGGACA
 1620
 F P L D A V L Y A D F S H D N T M T G I F S A M G L Y N G T
 AACCCGCTGCGACGCTCAAGATTCAGCCCTCCGACGGGTGCAAGGACGGATGGATACTGGGATCTGGGACGGTGGCTGGAGG
 1710
 K P L S T S K I Q P P T G A A A D G Y A A S W T V P F A A R
 GCGTATGCGGAGTTGCGATGTGAGACGGAACCGAGCTGAGGAGGGAGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 1800
 A Y V E L L R C E T E T S S E E E E E G E D E P F V R V L V
 AATGATCGGGTTGCGCTGATGTTGCGGGTGTGAGATGGGGAGGTGCGGAGGGATGAGTGGATTAGGGACTCACGTTGCT
 1890
 N D R V V P L H G C R V D R W G R C R R D E W I K G L T F A
 CGACAGGGTGGGCACTGGATGCGCTGCTTGTGATGAGATGCTCATAGACATAACCCATGATCCGAATTGATGTTTATAGATCA
 1980
 R Q G G H W D R C F .
 CTGGGAAAGGGAAATGATCCAAAAGGCCAGTCAGTATAACTTGCAGATCCGGTACTTGTCACTTGGTGTGCCATCAACC
 2070
 AGCCCTCCACAAAGCTCAATGTCCTCCCTCATGGAGTCCTGCGCCAGAGTCATCCACGCCAGGGCACGGAGCTGTTGGT
 2160
 AGGGTATCTGCCGTGTTGACCCCGTGCTCACAGTCACA

10	20	30
1 HAGIGLGSPLVLLLOPSAALLTASPAI P P P W		
1 M --- G V S A V L P E Y L L S G V - T S G L A V P A S R		
<hr/>		
40	50	60
11 R K X H P N V - - - - - D I A R H W G Q Y S P F F S		
27 N O S S C D T V D Q G Y Q C P S E T S H L W G Q Y A P F F S		
<hr/>		
70	80	90
52 L A E V S E I S P A V P K G C A V E P Q V L S R H G A R Y		
57 L A N E S V I S P E V P A G C R V T F A Q V L S R H G A R Y		
<hr/>		
100	110	120
82 P T A H K S E V T A S L L O R I Q D T A T E P K G D F A P L		
87 P T D S K G K Y S A M T Z E I Q N A T T P D G K Y A P L		
<hr/>		
130	140	150
112 R D Y A Y H L G A D N L T R P G E S O M M E S G R O F Y H R		
117 R T Y N T S L G A D D L T P P G E Q E L V N S G I K F Y G R		
<hr/>		
160	170	180
142 V R E Q A R E I V P F V R A A G S A R V I A S A E P F N R Q		
147 V E S L T R N I V P P T R S S G S S R V I A S G K K P I E G		
<hr/>		
190	200	210
172 F Q D A K D R D P R S N K D Q A E P V I N V I I S E E T G S		
177 F Q S T K L K D P R A Q F G Q S S P K D P V V I S E A S S S		
<hr/>		
220	230	240
202 M N T L D G L T C P A A E F A P I - D P T O P A E F I Q V F G		
207 M N T L D P G T C T V F E D S E L A D P V E A N E T A T F V		
<hr/>		
250	260	270
231 P R V L K K E T K H M P G V N L T L E D V P L P H D L C P F		
237 P S I R Q R L E N D L S G V T L T D T Z V T Y L N D N C S P		
<hr/>		
280	290	300
261 D T V G S D P V L F P R O L S P P C H L E F A D D W H A Y D		
267 D T I S T S T V - - D T K L S P P C D L F T H D E W I N Y D		
<hr/>		
310	320	330
291 Y Y Y T L D K Y Y S H G G G S A F G P S R G V G F V N E L I		
295 Y L O S L K K Y Y G H C A G C N P L G P T Q G V G Y A N E L I		
<hr/>		
340	350	360
321 A R M T G N D P V K D H T T V N H T L D D D N P E T F P L D A		
325 A R L T - H S P V H D D T S S N H T L D S S P A T F P L M S		
<hr/>		
370	380	390
351 V L Y A D F S H D N T H T G F P S A H G L Y N G T K P L S T		
354 T L Y A D F S H D N G T I S I L F A L G L Y N G T K P L S T		
<hr/>		
400	410	420
381 S K I Q P P T G A A A D G Y A A S W T V P P A A R A Y V E L		
384 T T V E N I T - - Q T D C P F S S A H T V P P A S R L Y V E N A		
<hr/>		
430	440	450
411 L R C E T E T S S E R E E E G E D E P F V R V L V N D R V V		
412 M Q C Q A E - - - - - - - - - - C E P L V R V L V N D R V V		
<hr/>		
460	470	480
441 P L H G C R V D R W G R C R R D E W I K G L T F A R Q G G H		
442 P L H G C P V D A L G R C T R D S P V R G C S P A R S C G G D		
<hr/>		
471 W D R C P		
472 W A E C F A		

Figure 3

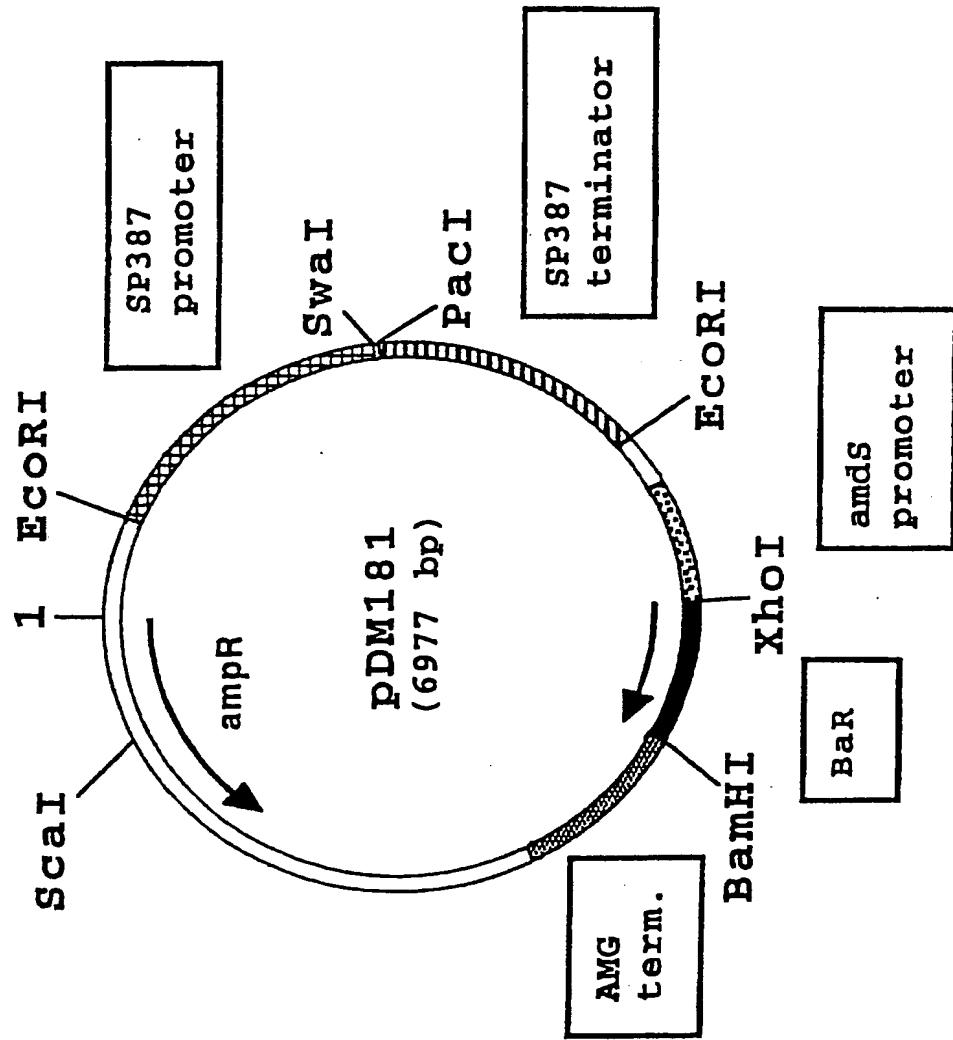
FIGURE 4

Figure 5

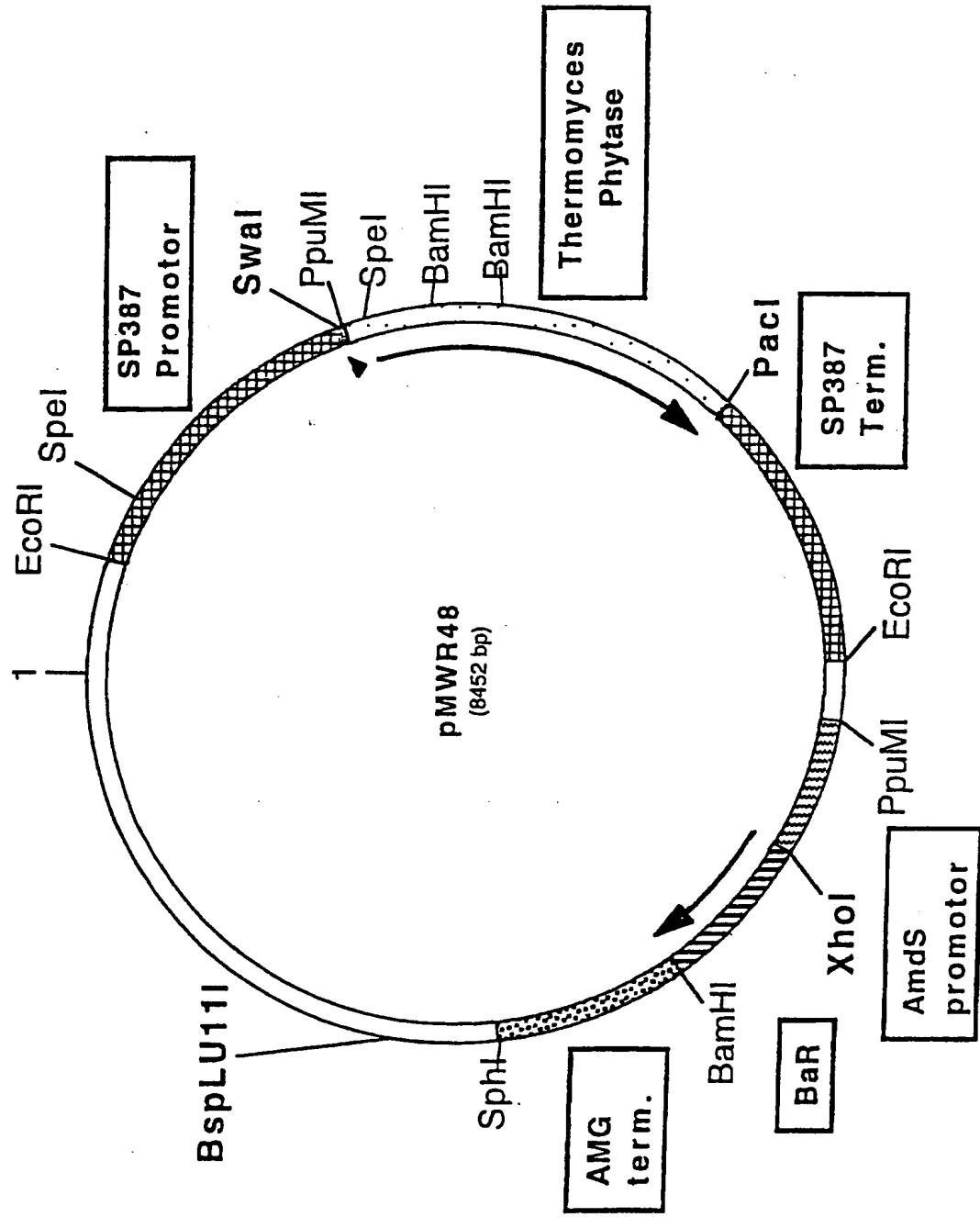


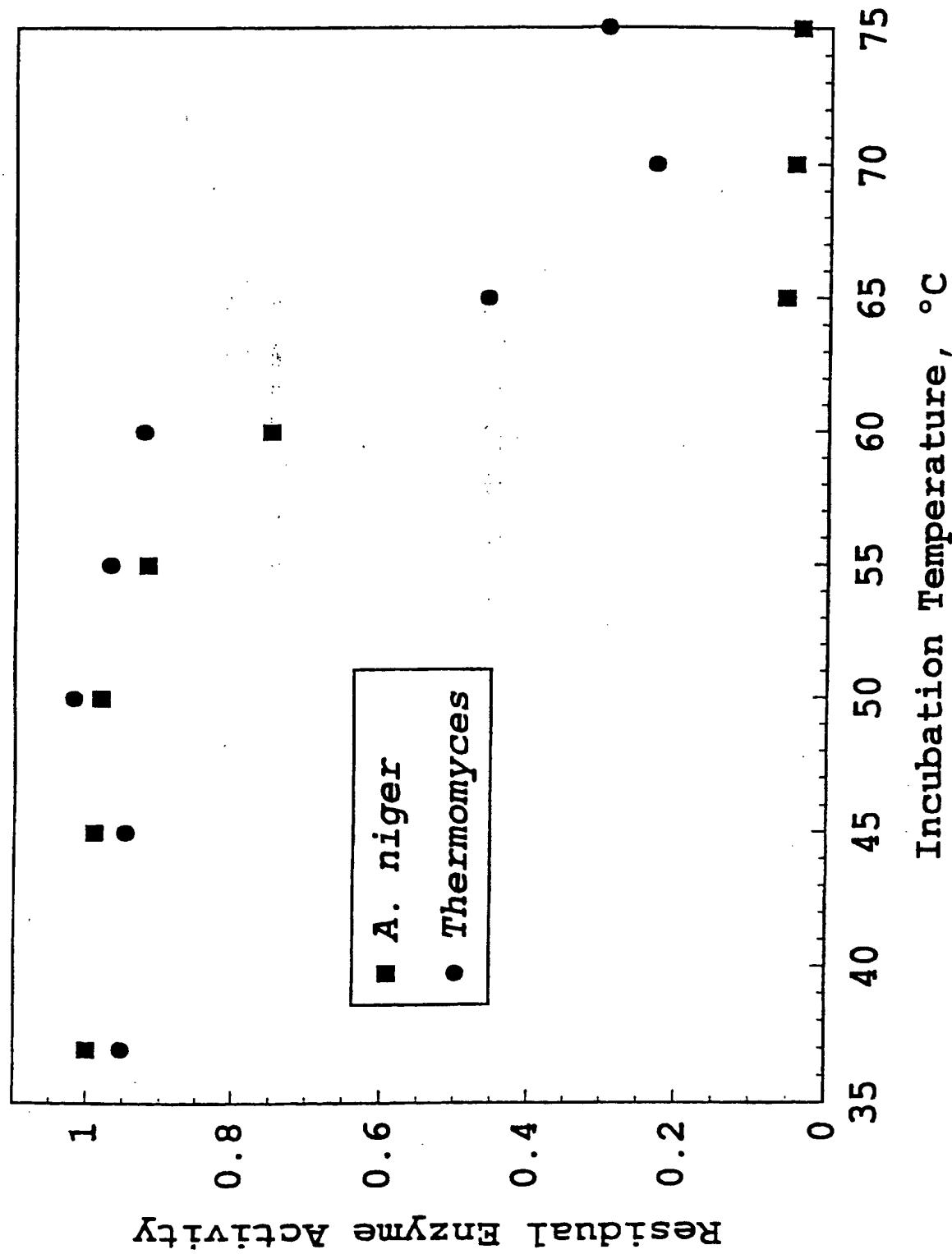
FIGURE 6

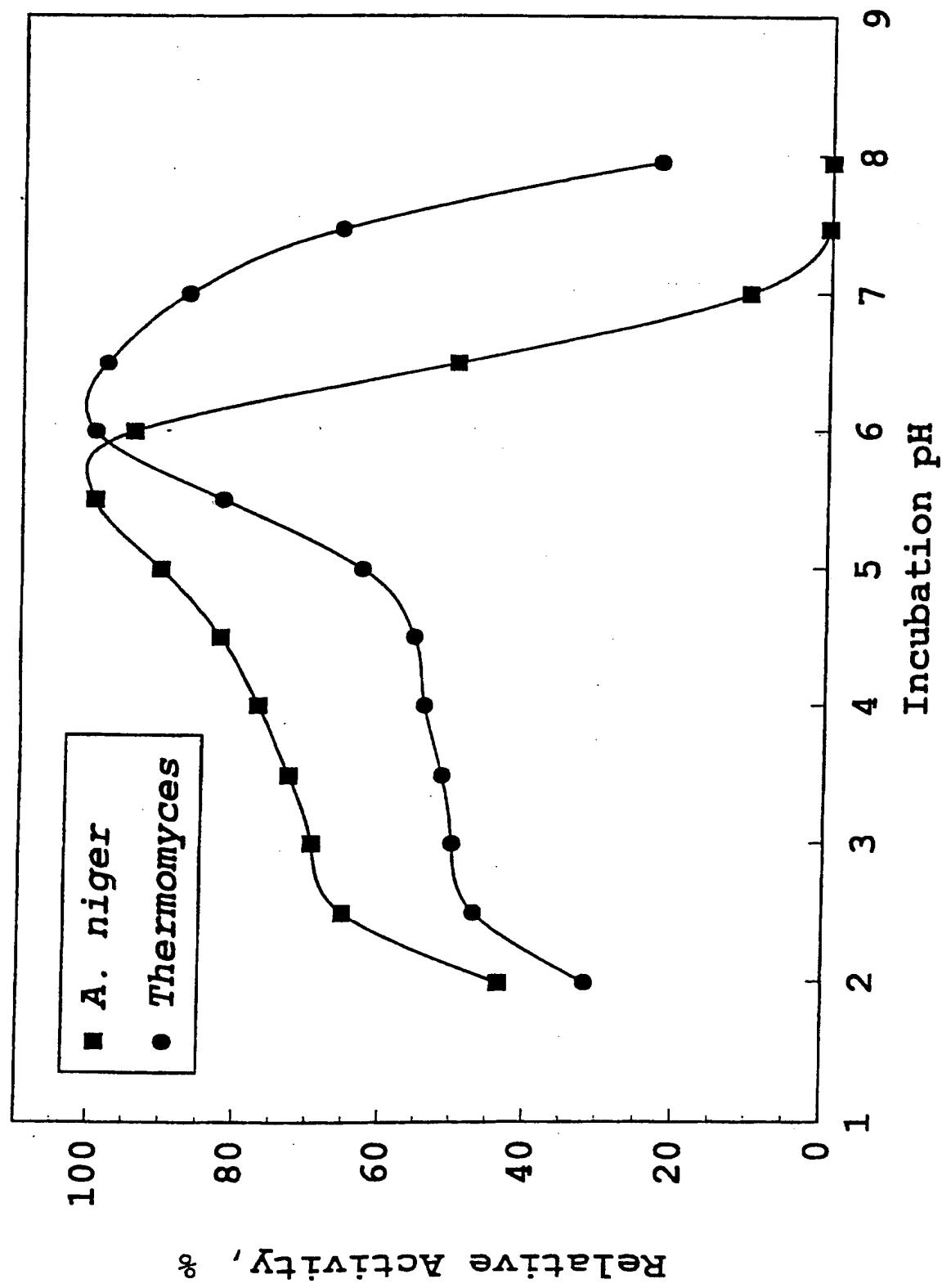
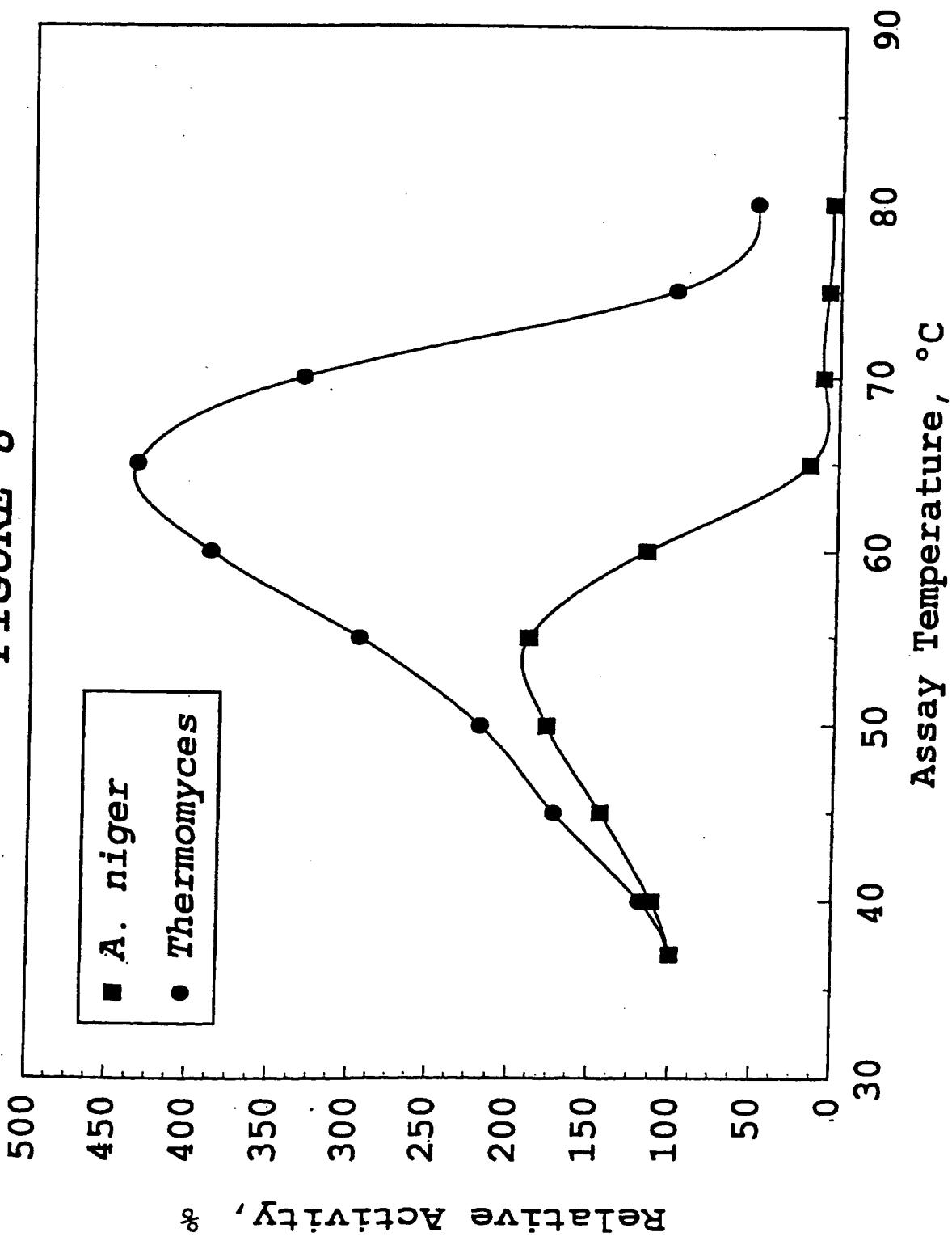
FIGURE 7

FIGURE 8



INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/03866

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/55	C12N9/16	C12N1/15	C12N1/21	A23K1/165
C12P19/14	//(C12N1/15,C12R1:77),(C12N1/21,C12R1:19)			

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A23K C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 420 358 A (GIST BROCADES NV) 3 April 1991 cited in the application	1,12,14, 18,23, 27-37, 39-41
Y	see page 1, line 1 - page 5, line 3 see page 27 - page 28; claims 1,22-31 ---	42
Y	JOURNAL OF FOOD SCIENCE, vol. 52, no. 3, 1 May 1987, pages 719-721, XP000574816 KNUCKLES B E ET AL: "EFFECT OF PHYTATE AND OTHER MYO-INOSITOL PHOSPHATE ESTERS ON ALPHA-AMYLASE DIGESTION OF STARCH" see page 719, left-hand column see page 721, left-hand column, line 28-42 ---	42

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

1

Date of the actual completion of the international search

21 July 1997

Date of mailing of the international search report

30.07.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+ 31-70) 340-3016

Authorized officer

Macchia, G

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 97/03866

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 96 28567 A (GENENCOR INT) 19 September 1996 see page 8, line 10-24 ---	1,42
A	EP 0 684 313 A (HOFFMANN LA ROCHE) 29 November 1995 see page 2, line 1-3 see page 2, line 53-56 see page 16, line 11-40; example 12 see page 17; table 3 see page 39-41; claims 1,20,23-25 -----	1-42

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 97/03866

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0420358 A	03-04-91	AU 636673 B AU 6501190 A BG 60108 A CA 2042054 A CN 1051058 A DE 420358 T EP 0779037 A ES 2072834 T JP 4506007 T LT 1527 A,B LV 10310 B WO 9105053 A PL 167790 B PL 168470 B US 5436156 A	06-05-93 28-04-91 15-10-93 28-03-91 01-05-91 12-10-95 18-06-97 01-08-95 22-10-92 26-06-95 20-10-95 18-04-91 30-11-95 29-02-96 25-07-95
-----	-----	-----	-----
WO 9628567 A	19-09-96	AU 5298496 A	02-10-96
-----	-----	-----	-----
EP 0684313 A	29-11-95	CN 1126243 A JP 8056676 A	10-07-96 05-03-96
-----	-----	-----	-----

